当代数学讲堂

DANGDAI SHUXUE JIANGTANG

数林奇葩

赵子都 著

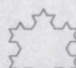
$g(x,y,\theta)=d+\cos\theta+\sqrt{(x-\cos\theta)^2+(y-\sin\theta)^2}$。

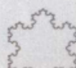
$F_n=\frac{\sqrt{5}}{5}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right]$。

$F(s)=\int_0^{+\infty} f(t)e^{-st}dt$。

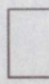

(a)

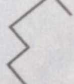

(b)

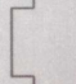

(c)

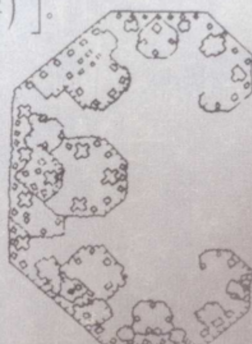

(d)


(a)


(b)


(c)


(d)


(e)

山东教育出版社



DANGDAI SHUXUE JIANGTANG

责任编辑：孙金栋　陆　炎
装帧设计：张　晶



ISBN 978-7-5328-6236-8

定价：37.00元

当代数学讲堂

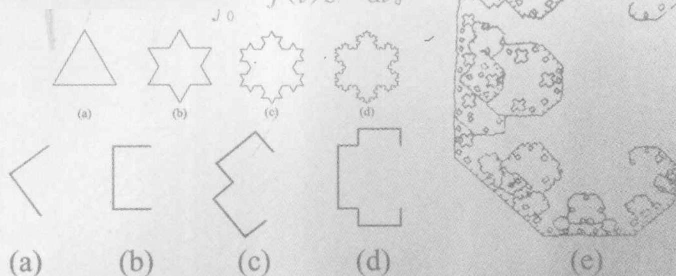
数林奇葩

赵子都 著

$l+\cos\theta+\sqrt{(x-\cos\theta)^2+(y-\sin\theta)^2}$。

$\left(\frac{1-\sqrt{5}}{2}\right)^n\Big]$。

$f(t)\mathrm{e}^{-st}\mathrm{d}t$。



山东教育出版社

图书在版编目(CIP)数据

数林奇葩/赵子都著. —济南:山东教育出版社,2009
ISBN 978－7－5328－6236－8

Ⅰ.数… Ⅱ.赵… Ⅲ.数学—文集 Ⅳ.01－53

中国版本图书馆 CIP 数据核字(2009)第 100212 号

那一代知识分子走了

——父亲、教授和数学家的故事

赵　瀛

1

有关人类"父亲"的角色在我的印象里却是淡的。他生前与子女的交流很少，家人也都习惯了他永远来去匆匆的脚步和深夜伏案的身影。9年前他去了另一处地方继续求解方程式，俗世的人们怎能不加以体谅：他的生命本属于事业，属于数学。

在生命的一呼一吸之间，父亲断断续续地对母亲说："我写的……谁还会看啊……"他为最后的数学留下了悲情的伤感和深情的眷恋。随后，那一缕凄美而微弱的、曾经绽放的数学光芒，随着一个注定时刻的到来而逐渐褪尽，悄然隐没于一个急功近利的年代。

父亲寂寞地走了。尽管一切按照既定的程式和俗例为他送行，但，那些不是他，也不像他，与他的生命品质并不接通，与数学的律动并不协调。在通常的人情伦理之内，我们有时也许并不了解身边最近的人，甚至可能轻易而慈悲地稀释某种价值。

20世纪80年代初，一位不速之客从太平洋彼岸的美

国来到中国辽宁，专程看望父亲。他们两人是大学同学，毕业后，一个去了台湾，一个留在了大陆。我一点也不奇怪父亲当时的选择。在国统区，他耳闻目睹的，是抗战结束后国民党“接收”大员疯狂地中饱私囊和贪污腐败，是社会的无道与百姓的疾苦，而新中国当是一个进步青年的希望所在！

时光荏苒，父亲申请入党 30 年，无论命运遭际怎样的坎坷曲折，都没能使他改变这个信念，直到 1982 年成为“党的知识分子”。纵观时势，有人入党是基于让更高的东西为自己服务，有人入党则基于让自己为更高的东西服务。后者被称之为信仰。

或许，他们两人当年是由于“信仰”的不同才分道扬镳的吧！

无疑，“信仰”这个抽象的词汇不是某种道具和高悬的理念，对父亲而言，它可以很具体、很平实，这恰与他本意要做的、自然而然要做的事情是一致的。

生前，父亲木讷，拙于表达自己，他完全不能够像表达和阐释数学那样表达和阐释自己。他的本性是一块质朴的、夯实的土地，坚守，甚至固执，但从没有一句怨言。今天，当人们仍然习惯于平面地了解事物，仍然热衷于眼前看得见的“发展”时，像这样一块块土地，作为一所大学最后的历史遗存的消失，究竟意味着什么?!

如果说，校园以学问作为安身立命之本，那么，对于一个诚实的知识分子，学术荣誉至高无上。一个学者、教授，只有也只能在他的脑力劳动产品中体现其对社会进步的贡献，进而体现其自身的价值，否则就无法将他拙扑的外表和精致的数学头脑统一起来。父亲去世后，我一直想却未能为他写一点儿什么。借助这本书的出版，终于拣起这支沉重的笔，将一直萦绕于脑海的种种线索加以梳理。值此写就的若干文字，并非要为他歌功颂德，乃是希望循着一个普通知识分子的生命轨迹，找寻属于数学和数学家的一种力量——它超越数学，但又无不在数学的每个方位折射出时代的镜像。

让记忆一一奔来眼底，历史深邃的目光穿越时间的雾障，希冀重新发现已被忘却掩埋的父亲和他曾拥有的缤纷数学世界。

2

1963 年的一个夏日，当异地任职的父亲走出火车站那栋黄色的日式建筑，这座辽西最大的小城以很高的规格迎接了他。市委组织部和统战部的领导向

他伸出温热的手掌，传递着厚重的政治寓意。父亲感动了，感动瞬间升华为崇高的使命感。他后来动情地对母亲说，心里真的有点不安，还没做什么事情，一定得努力工作，不辜负领导的期望。

父亲调任前在国家煤炭部所属的一所大学任教，事业发展正值健旺：28 岁当教学负责人，37 岁评定国家教授职级工资待遇，先后荣膺各项学术、政治荣誉，包括市劳动模范。母亲那里多年珍藏着他获得的数枚奖章。建国初期正是国家新兴的经济发展时期，那是一个单纯、上进和富于理想的年代。

新的学院建在城市北郊的高岗上，校园共有两栋楼（教学楼和宿舍楼）、一个小礼堂（兼食堂）、一排平房、一个篮球场和一座因供水困难而建起的水塔；学院的教职员工均住烧火炕的简易瓦房，前后的空地用秸杆围成个小院子，校园外的四周则除了荒地就是一望无际的高粱。如今想来，这里的条件不可谓不贫薄、不寒酸，好在当时的人们匮乏"物质文明"的概念，顺便也将春天的扬沙和冬季的北风一起接受了下来。印象中，还有来自上海、杭州等地的南方籍教师，还有来自名牌大学的毕业生，都陆续在这个寒冷而粗砺的北方小城安家立业。在这所由中专升格的、也许是全国最小的大学里，教师大部分从外地调入，各级领导多为工农兵干部，而来自全省的学生淳朴且非常好学。

大学虽小，毕竟具备了那个时代的大学所有的元素，并在上述背景下演绎了 50 年高等教育的曲折历程。

我的追溯应该说与这本书密切相关。书的作者以恒久的生命时间在这偏僻的一隅之地，耕耘着"党的教育事业"，也就是说，他来了，他们来了，就再也没有离开过！

父亲是一个典型的工作狂，工作起来全身心投入，且犹如数学解题，一丝不苟。他创建了基础部。不过这时政治风向已经越来越偏向"左"，数学也经常要和意识形态连在一起，有着山东人刚直不阿的禀性的父亲，在"被改造对象"的阴影下，难免遭遇掣肘。他不能在行政仕途上得心应手，却在专业领地如鱼得水。

教书或者受益于他的知识传授给了父亲很大的成就感。他常年承担繁重的教学任务，先后主讲高等数学、线性代数、概率论与数理统计、积分变换、矢量分析与场论、数理方程与特殊函数等课程。他对每一门科目都进行过深入研究，编有多种讲义和辅导材料。借助于广博的学识背景，他的课堂循循善诱，把抽象而枯燥的数学知识讲得有声有色。最难得的是几十年如一日，使之成为教学常态。

父亲深得学生的拥戴。我曾经与一些毕业多年的学生交谈，追忆当年的赵子都老师时，他们的言辞之间无不流溢出真诚的信赖和敬重……

父亲勤学苦读，志趣于术业，在 50 年代就开始了现代数学的研究。"大跃

进"时代的 1958 年，他发奋著述了 40 万字的《高等数学》(上、下册)一书，正待高等教育出版社出版之际，一场针对知识分子的"拔白旗"运动开始了。《高等数学》的作者被大加讨伐，指斥为只专不红的"一本书主义"，随即成为"一面白旗"，出版社被迫交还了已经排版的书稿。这是父亲经历的第一次打击，这个打击实在太残酷！父亲本是老实厚道之人，此时心惊胆寒，不知所措！幸好书稿没烧，后来学校印成讲义，成为在省内各大学通用的高数教材。这套讲义教材体现了父亲的学术造诣，也奠定了他坚实的理论研究基础。

待到"左"风吹到 1966 年"文革"时，知识分子人人自危，日子越发地不好过。其时的父亲成了被"打翻在地，再踏上一只脚"的牛鬼蛇神，脖子上挂着"资产阶级反动学术权威"的牌子站在被批斗的行列。不久，全国的校园卷起了大字报、大串联的革命狂潮，记得刚念初中的我也跟着跑出去"革命"了半个月，回到家中，却看到父亲端坐在书桌前，正埋头从事世界上最纯粹的逻辑思维活动……

在那些惊涛浊浪天天汹涌于耳根和眼前的日子，父亲纵然收容了一颗数学的心，也无力保住一张安静的书桌。几个月后，他被粗暴地剥夺教职，发配农村老家"改造"。

1968 年的一个黄昏，被遣送还乡的父亲一个人默默地走了，走向 5 年前热情迎接他的那个火车站。他几乎没有随身的行囊，手上只拎着一箱数学书稿。

是那箱数学书稿没有使他崩溃，就仿佛在绝望的挣扎中抓住了悬崖旁的一棵小树。他在承受的，许多人也都在承受。

他走向空旷的苍凉，心有多么大的苍凉啊，小城的季风在苍凉的天地之间呜呜地叫……

3

人，没有永远的逆境，也没有永远的顺境。

用一个数学方程式表示，即：逆境＋顺境＝完整的人生。

父亲亦然。

能够真正续接 50 年代的蓬勃事业的，是"文革"结束后的近 20 年。校园的紫丁香花开了又谢了，百人大学办成了千人大学、万人大学。一切都在快速改变，父亲也经受了时间的改变带给一个人全面、透彻的检验。他的数学重获尊重，他的敬业有口皆碑。

80年代和90年代是父亲在数学园地耕耘且收获最大的人生时段，他在繁重的教学工作之余，密切关注学科领域的最新发展。80年代初，他为年轻教师开设了数理统计、线性规划、最优控制等前沿的学科课程，此时大多数高校数学专业尚未开展相关领域的教学。1984年，由他主编并与同仁合作，共同撰写了42万字的《运筹学》，当时国内相关的著述还很少。为了争取出版，他们四处奔波，无奈终因校方和出版社的"经济"考量而成泡影。该书至今仍躺在档案室里悄无声息了20多年。1987年，父亲完成《突变理论及其应用》一书。突变理论是以拓扑学、奇点理论为主要工具，针对不连续现象（自然的和社会的）进行研究的新的数学分支，一时很受瞩目。这本书后来得到一位同事的资助得以出版。1994年，该书在台湾地区由亚东书局出版，持续发行了数年，可惜父亲直到去世也不知此事！

父亲退休后返聘当不挂名的教学顾问，为了帮助青年教师提高教学质量，每天坚持听两节课，风雨无阻。1990年，学院本科学生在高等数学全省统考中失利，他受命"救急"，协调主讲、助教老师全力以赴，终令成绩跃升，由原来的不合格率85%，达到合格率95%而名列前茅。

65岁离开教学岗位后，父亲便集中精力和时间，一头扎进数学的深海遨游，竭尽热爱学理的本份。那才是他的大快乐。

数学的精彩无须舞台、声音和色泽，只在论证环扣、推导纵深、公理瓜熟蒂落的时刻。

自然科学依赖数学的分支学科，已超出一个世纪的历史。一个有趣的现象是：数学知识的获得并不或很少依赖于其他知识，而其他知识的获得却要频繁地求助于数学知识。尤其在近20年，数学正突破传统的应用范围向几乎所有的人类科学知识领域渗透，化工、医疗、地矿、军事、农业、水利、电力、管理乃至经济、财贸、生物、环保、金融、信息技术等高新科技行业，都无一例外地运用数学的工具解决实际的问题。严谨、抽象的数学展示了其广泛的应用性，更加牢固地确立了科学技术的基础地位。父亲赶上了这样一个时代，一个数学和自然科学、社会科学的关系从来没有像今天这样密切的时代。数学，简言之，究竟能为人类认识客观世界提供什么样的知识和方法？将怎样为增进人类的福祉做出直接的贡献？类似的问题令他兴奋，令他敬畏，令他怀着孩童的好奇，探求那无穷的未知。

图书馆是父亲最殷勤的去处，他熟悉那里几乎所有的专业图书，一位老师说他"架上取书如囊中取物"。他的英文较好，可以直接阅读英文书刊，他熟悉著名的美国数学杂志《数学评论》就像熟悉中国数学学会主办的《应用数学学报》，他的书柜里也不乏英文原版的经典数学著作。没有国际的视野就没有数学的视野。一直以来，美国和西方发达国家的数学科学都处于世界领先地位，在国际上非常活跃的应用数学和新兴学科领域，数学家们在前人知识经验的基

础上，不断地创新着对数学科学和未来发展产生重要影响的成果。在国内，数学科学也日益受到重视，某些学科领域的研究达到了世界先进水平。

在令人振奋的全球化背景下，国内外许多具有开创性、奠基性的学说理论一经发布，热议的论题一经展开，父亲即投入研讨，兴致勃勃地撰文介绍。对近年来出现的边缘科学，他亦反应敏锐，兴趣盎然。早在 1979 年，他就开始研究模糊数学，接着又研究黑箱方法，不断有人上门请教他，大家都觉得新鲜而神秘。分数维几何刚问世，他就搜集资料，潜心钻研。他常说，数学是一切科学的钥匙，许多学科在使用数学工具后，才有了重大突破；许多问题通过数学的方法，才找到了科学的解答。有机会他还与同仁共同讨论，那时，平素寡言少语的父亲会滔滔不绝，神采奕奕，完全变成另一个人了！

计算机的诞生使人类进入信息时代。80 年代初，当计算机技术——这艘科学巨轮驶入中国刚刚开放的海域，父亲即着手研究计算机科学的核心课题。无疑，计算机的应用价值拥有无限的诱人前景，这个与数学紧密相连的学科，始终是父亲不断深入探究的兴趣领域之一。

中国在 14 世纪以前一直是世界上数学最为发达的国家之一，其以计算为中心、具有程序性和机械性的算法化数学模式源远流长，它们与古希腊以几何定理的演绎推理为特征的公理化数学模式相辉映，交替影响了世界数学的发展。父亲对中国最古老的数学名著《九章算术》津津乐道，深引为自豪，他的思维徜徉在智慧、敦朴的古典数学世界，细致地领略祖先开启数字奇迹的曼妙。

数学，如果是一种思想、方法和艺术，就不会仅仅意味着艰涩的公式和深奥的定理。用数学的视角探寻事物内部的结构，描述事物的规律和原理，在数学家们看来，不仅引人入胜，还充满着理性之美！数学，愈了解之深入，愈洞悉之神奇。这块神奇的天地令父亲的发散型思维纵横驰骋，探寻的触角伸进多个学科领域。混沌理论、分数维几何学、大系统理论、计算机算法、现代控制理论等都是近年来数学界的研究热点，父亲孜孜不倦，投以满腔的热忱，许多的阐发不乏独到的识见。在运筹学、突变理论等学科领域，他的探索也不断深入，绝非蜻蜓点水，浅尝辄止。总之，20 世纪数学各学科门类的发展气象万千，为父亲的研究范围之宽、涉猎之广提供了开阔而诱人的选题。

除了吃饭、睡觉，父亲整日坐在堆满文稿的书桌前写作。经常地，他同时思考不同的论题及其联系，同时撰写两篇不同内容的文稿。写到深处，家人唤他就餐几遍十几遍也不应声，他是舍不得离开迷人的推理乐园。有几次，他绞尽脑汁，终于求证出一道方程或破解了一道难题，竟欣喜若狂，连声嚷着："咳！我真笨，真笨啊！"父亲正是用无数的笨功夫擦出一次次顿悟的火花。

在生命的最后阶段，父亲共撰写了一百多万字的文稿，陆续刊载于《知识工

程》、《现代化》、《知识百科全书》、《自动化博览》、《自然辩证法》、《电脑爱好者》、《自然杂志》等国家级专业学术刊物上，有近60篇论文已储存在电子中文期刊数据库系统。晚年的父亲，其思维似乎比任何时期都富有活力。

同事们无不佩服父亲"能钻进去"，岂不知，他凡事钻进去就会陷进去，且越陷越深，很难拔得出来。他钻进教书里面，就全心全意陷进教书里面；钻进研究里面，就全心全意陷进研究里面。尽管身有疾病，尽管只有三分之一的胃（因"胃穿孔"，胃切除三分之二），只有二分之一正常的躯体（因脑血栓，一侧行动不便），他却从不把自己当成病人。其实，他在去世前的一两年经常因病住院，可是只要一出院，就照例回到书桌前。这已成为一种习惯，让他的大脑不停歇地运转，直到拿不动一支笔。1999年，他的最后一篇论文《遥感、遥测和遥控》写成后，就再也没有走出医院。那一年他75岁。

那一次，我回到家中，父亲迫不及待地拿出新近发表的论文给我看，神色疲惫的面容洋溢着孩子般单纯的、快乐的光彩。

有多少人愿意理解这快乐的涵义呢？他的书斋是寂寥的。没有同道，没有评奖，没有人要求，没有人与他分享思考的愉悦。在日复一日的狭窄的书案前，一个最简单、最美妙的数学证明就是他的节日。

当商业文化的浮华之气不断浸润学术大环境时，远离功利的基础学科总是显得默默无闻，以至当经济学家们在无数个耀眼的论坛上叱咤风云的时候，大多数的数学家们依然远离公众视线，他们的工作似乎是"无用"的，尽管"他们所做的事情可能是渺小的，但这些事情却具有永恒的性质"。

科学家们认为，数学是各门科学在高度发展中所达到的最高形式的一门科学，是人类文化发展到高度智慧的标志；没有相当发达的数学，就没有相当发达的科技。因而，数字化时代需要更多的数学志士追求数学的真理，而数学的真理归根结蒂是求真的态度，是求索的精神。

在那一瞬间，我读懂了父亲对数学的痴情……

4

1995年，在昔日的一位学生的帮助下，父亲将部分发表的论文印成一本书。严格地说，那只是一本外表极简陋的铅印的小册子，但父亲心满意足。当他用双手捧着这本"印刷品"赠给朋友时，就像捧着一件无比珍视的藏品，满脸庄重。

那时，他当然不会想到，在他身后，此时遥远的故乡正在解读他和他的数学，并为一个普通知识分子的快乐的价值所打动！

1994年，山东教育出版社设立了全国首家资助教师专著出版的教师出版基金，她高远的立意令我期待。通过基金办公室陆炎女士的热诚努力，终得以将父亲的小册子整理出版。我由衷地感谢山东教育出版社对父亲和他的数学给予的深切理解。让知识超越市场，他们无疑做了一件有意义的事情，而在整个出版界中的这样一种超越，总会令人肃然起敬。

我以《数林奇葩》为父亲的这本著述命名，该书遴选数学论文65篇，凝聚了父亲晚年的全部心血。综观内容文字，虽然专业性很强，但深入浅出，表达流畅，颇有一点讲座的风格。是的，这里还是一个当代数学讲堂，讲演者站在讲台上，熟练地使用他的数学语言，或旁征博引，或精确刻画，而他的听众则是专业人士、莘莘学子以及广大数学爱好者，他们将从不同的层面获得数学的启示。

随着各种职业对数学的需要，经济师、软件设计师、统计师、工程预算专家、网络安全专家等越来越依赖数学知识开展自己的工作，那么该书当为其了解相关数学知识提供有意义的参考。对于年轻的学子，《数林奇葩》是提升数学修养的简约读物，是专业知识的补充和丰富。它犹如数学森林里几十棵类型各异、富有活力的植株，让求知的心灵因呼吸、触摸它们而视界开朗，灵思蓬勃。中科院院士、著名数学家杨乐认为，一个人的分析、思考、计算以及逻辑推理能力往往决定其创造力和专业水准，他称数学是培养科技领域高素质人才和创新人才必不可少的基础。

近期查阅相关信息，得知父亲的一些论文先后被国内的大学、科研机构研究借鉴。一位清华大学毕业的年轻教师说，他在撰写硕士论文时，导师为其指定的参考资料就有赵子都教授的论文。一般地说，论文的被引用率是考量其学术价值的重要参数。看来，父亲晚年的研究成果正在引起学术界的注意。我常想，如果父亲的学术之帆停泊在大一些的港湾，也许会在更宽容的空间发挥潜能，那么他的冲浪一定会更精彩！

没有人想到，父亲原本应该成为一名律师。

少年时代的父亲数理成绩就总是班上的第一名，高中毕业后曾以优异的成绩考入上海交通大学，因无力支付昂贵的学费，转而考入国立武汉大学文科。虽然当时社会重文轻理，在珞珈山的学业之余，他还是经常旁听数学，自学了全部的高数课程。1950年，已经取得在一所学校教授数学资格的父亲经同学同乡、现中国工程院院士、东北大学教授张嗣瀛伯伯的引荐，来到百业待兴的东北。凭着勤奋努力，他的学术成长很快，学术势能不断积累。在“文革”中，有人揭发他是学法律的，不懂数学，引得一时哗然。他既然不会为自己辩解，就更加

惶恐于自己的“出身”了。

不过这件事倒很巧合——我指的是法国人费马(Pierre de Fermat),他也是学法律专业的,却是世界公认的大数学家。费马当了34年公务员,而卓越的数学贡献都是在这个期间成就的。费马有一句名言:“数学,学者业余的宗师。”当然还有更多类似的佳例。无论是跨学科的复合型人才还是某个方面具有禀赋的人才,不过是人类一种规律性的智力现象,并非鲜见。倒是面对文理两个专业之间的巨大鸿沟,能够如此从容地跨过去,其中是非付出几倍于常人之努力不可的!费马是天才,父亲不是,父亲常说自己是个笨人,共通之处只在于冥冥中一只造物主的手将他们推向一个别无选择的方向:热爱。

5

对事业有多高的标准,对生活就有多低的标准,这是一个只适于父亲的逻辑推演。在物质极大丰富的今天,父亲对物质生活却没有任何要求。他强烈反对装修房子,家人谁要提及装修,他就立刻和谁急;他拒绝为他购置任何物品,万不得已时,又坚决要求廉价实用;他几乎从不消费,从不去商店,偶尔与母亲同去,也必定坐在外面的台阶上等待。他对自己苛刻有加:虽然喜欢吃香蕉,但若1斤3元以上的价格,他会远远地绕开;他与同事出差,为了节省交通费,宁可在骄阳下步行。公私在他那里是泾渭分明的:经常使用的稿纸和墨水,从来都是用自己的钱购买;常年取自单位的《参考消息》都是个人订阅的……

在世人看来,他真是“呆”得可爱呀!

母亲说他艰苦奋斗一辈子。有关他“艰苦奋斗”的故事可随手拈来。3年困难时期,他扛着锄头去开荒种地,经常披星戴月。有时他带着作为家中长女的我做他的帮手,分配给我的任务是捡拾荒地上数不清的小石头,但我常常捡了一会儿就烦了,对着不远处那四处堆起的小山包似的石块发愣:那是父亲用最原始的劳动垒筑起来的“样板工程”。每次去地里干活儿,父亲都会引我去“参观”贫瘠的土壤上一片片长势不良的庄稼,“这是玉米!那是毛豆!这是荞麦!那是芝麻!……”他兴冲冲地、一一地指点着,如数家珍。“荞麦!芝麻!老家种的!”他高声地强调着,显得特别开心。他希望有人欣赏他的劳动成果,可遗憾的是,我们却从没有收获过这些好东西。那时我还是个小儿童,受着父亲的经常派遣,一会儿扒煤渣儿,一会儿拾白菜叶儿,忙得不亦乐乎。后来由于母亲

的激烈反对，他才略加收敛。

父亲为人忠厚，他许诺于他人的事情，一定要抓紧兑现，心里才安宁。对他人嘱托或指示，则字字相信，回回照办，不肯一点儿敷衍，母亲常笑他“拿棒槌当‘针’（真）”。他素来极少求人，唯恐给人家带来麻烦，他与张嗣瀛伯伯是挚友，惟有的两次求助，却都是为了帮助学生考取研究生。他做人正派，是非分明，遇到悖于合理之事就直言不讳，自己则表里如一，从不谋私利。他一贯主张开短会，不搞形式，不扯咸淡。他躬身践行“实事求是”，一生从不说假话、大话、谎话、空话。

“取”字是一个敏感的词，对此，父亲有严格定义：“取”，即向自己索取。他一生所成，皆因自己的辛勤，决不窥测身外捷径。他竭尽努力，最不吝身心的付出，即使是没有结果的付出。在各种利益面前，“取”字堪为对人性的最大考验，偏偏父亲最显天生拙笨，全无“取来”之心机，一辈子多吃眼前亏，不知有便宜事。

“呆”的父亲，却自视很高，一向只服膺和推崇站得更高的人：吴文俊、华罗庚、陈省身、苏步青、陈景润、熊庆来……还有那些熠熠生辉的世界著名大数学家的名字。当他仰望头顶的星空，却忽略了脚下的生存技能。他的数学世界与现实生态的种种潜规则格格不入：待人处世缺少弹性，拙于交际，拙于家政，拙于处理“关系”之类的人事纷繁。生活的“弱智”倾向令他在数学以外的表现“乏善可陈”。在通常的人生经验里，这个教授显得“一无用处”。的确，对于父亲来说，那些棘手的事情都不是数学所能解决的。

性格也属于非智力因素。有时他脾气暴躁令人难以容忍，母亲往往迁就他，知道“暴风雨”来得快，去得也快。由母亲撑起的“后方”屋宇，是父亲的数学精神得以滋养和安宁的伞。

如此，一件事物的两面性在父亲的个性里，恰如高山峡谷般地突兀而呈现出令人诧异的反差，似乎造物主在给予某人一些优点的同时，又必须给予一些对应的缺点，将强和弱、智和愚组合在同一个人身上。由于人的多元性，社会心理当能容下在某个方面过于专注和在某个方面过于疏忽的个人方式。然而人性的不完美并非是可以原谅不完美的理由，因此不能否认父亲性格的某种缺失所导致的非智力因素的缺失，对命运机制产生的不利影响。

6

父亲的家乡是个依山傍水的大村庄，有沙有土，适于各种农作物生长，所以

他的“开荒实验”可能源于植根在脑海里的童年时代的劳动图景。父亲对家乡一往情深，有谁论及泰山，谈及花生和煎饼，他必侃侃而谈，眉飞色舞。他一向对军事感兴趣，最愿意看抗日战争题材的影视片，《莱芜战役》、《台儿庄大捷》在电视上播映时，他兴奋得大呼小叫：“快来看啊！有好电影了！”嘴里还不停地提示着：“莱芜战役，家乡的战役啊！”然后，正襟端坐，全神贯注，一直看到完仍意犹未尽。

心情好的时候，父亲喜欢唱抗日时代的流行歌曲。“大刀，向鬼子们的头上砍去！”小时候，父亲教我们学唱《大刀进行曲》，但是谁都不愿跟他学。“同学们，大家起来！肩负起天下的兴亡……”那首《毕业歌》，父亲唱起来声情并茂，十分地投入，一支歌从头唱到尾，一字不落，唱到动情处，热泪盈眶……我们不知道，当年，一个瘦弱的少年站在饥寒交迫的流亡队伍里，他破衣烂衫，满腔悲愤，以热血之躯一遍遍地吼出国破家亡的切肤之痛！我们不知道，多年后，那一种强烈的情感还在他的血管里奔腾，那荡气回肠的旋律还在他记忆的深处盘桓……

父亲14岁离家到省城求学。1937年，日寇铁蹄长驱直入，齐鲁山河喋血，济南危在旦夕，不甘做亡国奴的师生们经由教育部门的组织集结，向内陆省份转移。正在济南求学的父亲跟随学校踏上了悲壮的流亡之路。从济南到四川，数千学生仅凭着一双铁脚板徒步行走，一路险渡黄河，翻越秦岭，行程一万余里，历时七个多月，陆续抵达四川绵阳。这段刻骨铭心的经历给父亲的一生刻下了深深的印记，成为他命运沉浮的第一个重大事件。他曾撰文回忆，当时家乡成了敌占区，家园一片焦土，师生群情激愤，抗日热情高涨，他和一些同学报考了黄埔军校，准备日后奔赴抗日战场。不过最终在老师的建议下，喜欢读书的他还是选择了跟随学校迁徙大后方。

设在绵阳的山东国立六中是中国教育史上浓墨重彩的一页，它是山东沦陷后唯一保存下来的一所完全中学，抗战时期向全国输送了大量的优良学子，功名卓著而赢得各界赞誉。父亲在这里完成初中学业，即以各科课程全优的成绩免试升入高中。当时前方打仗，后方供给困难，学校的学习生活条件之艰苦非常人所能想象。父亲说，在中学的六年里，他从未穿过鞋和袜子，就靠手编的草鞋赤脚度过了一个又一个寒暑。直到升入大学，他才用奖学金第一次为自己买了一双布鞋、一双棉袜。生活在无任何经济来源的窘况下，他想方设法给人家当校对，当兼课老师，每日黎明即起，夜深方归。父亲是家族读书人中勤工俭学第一人。

青少年时期的艰辛倍尝，磨砺出父亲的韧性，培植了他吃苦耐劳的特质，而难能可贵的是：他一生保持了这种特质，一生拒绝安逸。

7

父亲走进天国的时候，城市的清晨还未醒来。天国的门是洞开的，欢迎来自人间的一颗清洁的灵魂。

父亲从我们身边走过，以一生不变的步履走过。他以赤子之心投报这世界，并不计较这世界回馈给他更多的是坎坷、困顿，是逆境多于顺境。当时代穿行于物欲横流的喧腾中，他亦本色来，本色去，在随风潜人夜般的种种诱惑中坚持和自守。他恰如一股透明清新的泉水，只管潺潺流淌而无人觉察。

诚然，他位卑而无名，没有头衔，甚至从来没有名片。诚然，还有成百上千的数学家在辛勤地劳动，在贡献毕生的精力，而他只是他们中的一员。诚然，他的跋涉可能永远达不到辉煌的真理的终点，只是一路上发现了一些风景，捡拾了一些有兴味的片段。不，这些都不重要。

父亲、教授和一个数学家的故事如此告诉我们：一个人将怎样尽力做最好的自己，做应有的自己，直到生命的落幕。

每个时代都要经历特定的社会洗礼，每个人都会留下时代的印记。父亲的时代赋予了那一代人一些比较普遍的特征，诸如理想、追求、诚实、俭朴、勤勉、刻苦、踏实、专注、严谨、坚毅、执著、责任感……它们使那一代知识分子散发着一种特有的精神气息。如果说它们是一种品质，那么这些品质只不过是基本的人格修养，绝非惊人耀世。时代变了，但是作为一种常识的价值取向是不会变的，为此，没有人否认当今它们仍然是为教为学所需要的，是各种职业岗位所需要的。令人忧虑的是，在财富和器物价值驾驭人心的情势下，它们也许正在校园里流失，正在更广泛的社会范围里淡化，正在成为现代人比较普遍的稀缺。

常识的力量能够变成非常的力量，这应该是父亲那一代人留下的遗产，他们的遗产能否传承下去，他们的学养能否留驻下来，这并非不是一个值得面对的问题。

那一代知识分子走了，在四季迭换的目送里，走远了……

2009 年 3 月 23 日于北京寓所

Qian Yan

前言*

众所周知，数学是研究现实空间中的数量关系和空间形式的科学，简言之，是研究数与形的科学。由于数和形贯穿在一切科学领域，因此，数学的应用也就越来越渗透到各个领域，成为人们认识和驾驭客观事物的有力武器。在现代任何一门科学中，如果不借助数学，不与数学发生关系，便不能刻画客观事物变化的状态，更不能从已知数据推出未知数据，因而减少了预见的可能性，或削弱了科学预见的确切性。

在某种意义上说，数学是科学之母，是通向一切科学的钥匙。如物理、化学、计算机、金融越来越需要较多的数学知识，现代控制理论更是建立在精确的数学模型的基础上，乃至过去以描述为主的生物学、经济学也处在日益“数学化”的过程中。正如马克思早就预见性地指出：“一种科学只有成功地运用数学时，才算达到了真正完善。”

20 世纪 40 年代出现的电子计算机正在改变整个数学的面貌。电子计算机不仅具有高速运算能力，还具有逻辑推理功能。某些或某类定理的证明完全可以交由计算机来完成，从而使数学家得以从琐碎而又十分单调的加减乘除的劳动中解放出来。换言之，把质的困难转化为量的复

* 此文系作者于 1995 年所编著《数学、计算机与对策论》的前言。

杂，而后者对计算机来说轻而易举，使定理的证明化难为易。如悬置了一百多年的地图着色的“四色猜想”就是在1976年由美国科学家在计算机上成功证明，也因此使“四色猜想”变成“四色定理”。

离休以后，我有4年时间继续做教学方面的指导工作，以后时间就完全属于自己了，这使我得以全部精力和时间投入相关研究。围绕着上述的论题，我的论文刊登在国内有关学术杂志上，这本集子筛选了其中的一部分内容。科技的进步推动着数学这门基础学科深远拓展，当代数学已经同科技一起站到了时代的前沿，我时常为之兴奋不已，十分乐于和我的业界同仁及所有爱好数学科学的人们分享和交流。

我的生命已经很有限了，这是我向它贡献的最后一点东西。如果能给读者一点启示，尤其是能够给青年人以切实的帮助，我亦十分地欣慰了。

赵子都

1995年1月

Mu Lu

目录

数学：一种思想、方法和艺术

1 《九章算术》及其对现代数学的影响*

中国科学院院士、著名数学家吴文俊于80年代初首创了“机器证明理论”，该理论使数学机械化成为现实。国际学术界对此给予高度重视与评价，称其为“吴方法”，认为这是在“近十年中自动推理领域出现的最为激动人心的突破性进展，在自动推理领域具有划时代的意义”。这是我国一项继华罗庚“优选法”之后，对社会经济及科研有重大影响的方法。1990年，在北京“中国数学机械化中心”成立大会上，吴文俊解释说，他的这一方法直接导源于我国的数学思维方式，也就是从公元前1世纪成型的《九章算术》开始，经南朝宋齐时期大数学家祖冲之，到元代数学家朱世杰等形成的以解方程为特色的机械化算法体系。

《九章算术》是举世公认的一部最古老的经典数学名著之一，又是《算经十书》①中内容最丰富和最重要的一部，历来被列为算经之首。该书既非出自一人手笔，也不是一个年代的作品，其中有些是秦汉以前流传下来的老问题，后经历代名家如西汉张苍、耿寿昌等人删补增订，在公元前50年左右已成定本。到了三国时期，数学家刘徽作《九章算术注》，加上自己的心得，润色一番，使之便于理解，因而流传下来。刘徽在《九章算术注》的原序里说：“按周公制礼而有九数，九数之流，则《九章》是矣。”中国数学史家钱宝琮说：“九章的名称无疑是《周礼》九数的演变。”一般

---

*本文原载于《辽宁工学院学报》1992年第2期。

① 十书是：《周髀算经》、《九章算术》、《海岛算经》、《孙子算经》、《张丘建算经》、《五曹算经》、《五经算术》、《缉古算经》、《数术记遗》、《夏侯阳算经》。

认为，因为它将过去的246个数学问题分为九章，所以叫《九章算术》。

数学是研究现实世界中数量关系与空间形式的科学。《九章算术》中的246个数学问题，就其性质来说也都是研究"数"与"形"的。如第一章"方田"，是研究平面图形面积的，着重指出了分数的约分、通分、相加、相减和乘除法则；第二章"粟米"，主要是讲粮食交换的计算；第三章"衰分"，讲许多不同类型的比例（"衰分"有定量分配之意）；第四章"少广"，讲从田亩（平面图形）的面积或球的体积求其边长或径长的算法，在这章里出现了世界上最早的多位数开平方、开立方的法则；第五章"商功"，讲各种土木工程土方（立体）的计算；第六章"均输"（汉初实行均输法，是征收实物地租的法律），讲处理粮食运输、均匀负担等问题，所用的方法有配分比例、复比例、等差数列等；第七章"盈不足"，讲计算盈亏类问题的技术；第八章"方程"，讲联立一次方程组（线性方程组），也给出了联立方程组的普遍解法（见下文）；第九章"勾股"，讨论用勾股定理解决应用问题的方法。经深入研究发现，在中国传统的数学中，数量关系与空间形式是形影不离、并肩发展着的，在研究空间形式时也往往通过数量关系来处理，使几何问题转化成代数问题，即化几何问题为代数方程，然后通过方程的研究来探讨几何的属性。17世纪法国数学家笛卡儿（Rene Descartes）创建的解析几何学，正是中国古代数学的这种思想与方法在几百年停顿之后的重现和继续。吴文俊精辟地指出，中国的传统数学有它自己的体系与形式，有着它自身的发展途径与独到的思想体系，即基本上遵循了从生产实践中提炼出问题，经过分析与综合，形成概念，并上升到理论，提炼出极少数一般性原理，进一步应用于多种多样的不同问题。从问题而不是从公理出发，以解决问题而不是以推理论证为主旨，这与西方以欧几里得几何为代表的所谓演绎体系的旨趣迥异，途径亦殊。《九章算术》正是遵循这一途径与方法而发展起来的。

《九章算术》的特点之一就是全书以问的形式从社会生产和生活中提出了246个数学问题，然后再通过一个具体问题去说明一种法则（算法），并给出答案。然而，这并不失去算法的普遍性，因为这实质上并不是对一个问题的解法，而是这一类问题的解法。换句话说，可按这种算法去解决性质相同的问题。如果把用以解决问题的法则或算法称为"术"，那么，《九章算术》的基本方法就可以说是"由问而术"。所谓"问"即问题，是从生产实践中提炼出来的实际数学问题。每一章先提出一个或几个实际问题，并给出算法和答案，然后再考察同类型的问题而归纳出解决此类问题的一般算法（术），这便构成一章。最后再把解决各类问题的方法综合起来而成《九章算术》。故《九章算术》源于实践又高于实践，以解决实际问题为目的，有些算法实际上就是一种数学模型。这些一般算法就是《九章算术》的主要内容及其精华所在。以第八章"方程"中的第1题

为例："今有上禾三秉，中禾二秉，下禾一秉，实三十九斗；上禾二秉，中禾三秉，下禾一秉，实三十四斗；上禾一秉，中禾二秉，下禾三秉，实二十六斗。问上、中、下禾实一秉各几何？答曰：上禾一秉九斗四分斗之一；中禾一秉四斗四分斗之一；下禾一秉二斗四分斗之三。"

"禾"是黍米谷子，一"秉"是一束(捆)谷个子，"实"是打下来的黍米谷子。上述问题，用现代汉语来说，第一条就是：3 束上等的谷个子，2 束中等的谷个子，1 束下等的谷个子，打下来的黍米谷子共有 39 斗。其他各条也同样解释。很明显，这个问题实即一个三元一次方程组(线性方程组)。用现在代数的方法来解：设上、中、下秉的实各为 x,y,z，于是，依题意可列出以下方程组：

$$\begin{cases}3x+2y+z=39,\\2x+3y+z=34,\\x+2y+3z=26。\end{cases}$$

古代未知数不是用符号表示，而是用"筹"来表示的。"筹"原是用竹子制成的一种计算工具，有时也表运算，所以又称"算筹"。现用算筹表示 x,y,z 的系数和常数项，依次(自上而下，自右至左)布置起来，就得到一个如下所示的图表。

左 行	中 行	右 行
𝍠	𝍡	𝍢
𝍡	𝍢	𝍡
𝍢	𝍠	𝍠
𝍪𝍥	𝍫𝍣	𝍫𝍨
③	②	①

表内是古代用算筹表示的数目，如右行自上而下表示 3,2,1,39。有几个未知量就需列出几个等式。联立一次方程组用算筹表示时，犹如方阵，所以叫做方程。不过古代数学中的方程与现在的方程在含义上已大相径庭。现在方程是由英文"equation"直译而来，有相等之意。通俗地说，方程就是含有未知量的等式。《九章算术》中，算筹用以表示一个算式中的系数，这实际上就是现在代数中的分离系数法，然后用消元法求解。我们知道，这是解联立方程组的普遍而有效的方法。除了符号和计算工具不同外，和现代解联立方程组使用的消元法实质是一样的。由于形形色色的问题往往归结为方程求解，因而方程求解就成为中国传统数学发展的一条主线。《九章算术》不仅提出了线性方程组的普

遍解法——消元法，且由于方程的每一行是由多个未知量和一个已知量所组成的等式，其中可能含有相反意义的数，由此产生正数与负数的对立概念。又用"直除"法消元，当减数大于被减数时，不引入负数的概念就行不通了。因此，中国数学家在"方程"章里提出了正、负数的不同表示法和正、负数的加减法则。如"方程"章第3题刘徽注提到用红色算筹表示正数，用黑色算筹表示负数；或者在布置算筹时用正列的筹表示正数，用斜列的筹表示负数。这一章里还有四元和五元的方程，也用类似的方法来解。

其他国家虽然有时也解出一些联立方程组，但用语含混，不仅时间上晚于《九章算术》，且远不及《九章算术》算法整齐，也没有一般的解法。如希腊时代被誉为代数学鼻祖的丢番图(Diophantus)常常仅用一个记号来代表许多未知数。法国在1484年才由舒开(Nicolas Chuquet)给出二次方程的一个负根。意大利的卡当(Cardan)在1545年区分了正、负数，把正数叫做"真数"，负数叫做"假数"，并正式承认了负根。至于线性方程组的完整解法，到17世纪，德国数学家莱布尼茨(Leibniz)才着手研究，并由此出现行列式的概念。1764年，法国数学家艾蒂·贝祖(Étienne Bézout)用行列式去建立线性方程组的一般理论。从时间上说，《九章算术》关于线性方程组的解法及由此相伴而生的正、负数概念不仅是世界数学史上的光辉成就，也是科学史上破天荒的一件大事。

《九章算术》从实际问题出发，由简到繁，从特殊到一般，由具体到抽象，以解决实际问题为主旨的演变过程，不仅为此后一千多年来的数学，特别是方程的发展开启先河，而且在发展过程中，建立了以构造性与机械化为特色的算法体系。这与西方数学以欧几里得(Euclid)《几何原本》为代表的所谓公理化演绎体系正好东西遥遥相对，相互辉映。如果我们深入研究就不难发现，中国古代数学基本上是一种机械化数学。《九章算术》中对方程、开平方、开立方的机械化过程都有许多说明。如在开平方法中是借用一根算筹来表示未知量的平方，在开立方法中借用一根算筹来表示未知量的立方，这就给所列出的筹式一个代数方程的意义。开平方或开立方的各个演算步骤也就是解方程、求正根的过程。中国古代数学开平方、开立方的方法，不仅具有算术上的意义，更重要的是它们具有代数方面的意义。后世数学家求高次方程正根的方法无疑是在《九章算术》中"少广"章开方法的基础上发展起来的。在宋元时代，中国数学家们终于找到了一种普遍的列方程的方法——天元术，引进了相当于现代未知数概念的天元、地元、人元、物元，把许多问题转化成代数方程或方程组的求解方法，这个方法用在几何上就相当于几何的代数化。12世纪，杨辉、李冶、朱世杰又引进了相当于现代多项式的概念，并建立了多项式运算法则和消元法有关的代数工具，使几何代数化的方法得到了充分的发展。

几何的代数化实际上就是现代解析几何的前身。我们知道，17世纪笛卡儿发明的解析几何，虽然能用坐标法把几何问题化为代数问题，但并未实现几何定理的机械化（自动化）证明。吴文俊在深入研究中国古代数学，特别是在研究《九章算术》的基础上，在中国古代数学的机械化与代数化的光辉思想启迪下，独辟蹊径，利用最先进的、具有推理判断功能和存储功能的现代化计算机器——电子计算机，把质的困难转化成量的复杂，提出了自己独具特色的机械化判定方法——定理机器证明方法。□

2 古老而又年轻的二进制*

在电子数字计算机中，数是采用二进位制计数法（简称二进制）表示的。

就数的进位制而言，我们通常使用得最广泛的是十进制，其中任意一个数是由 0，1，2，3，4，5，6，7，8，9 这十个数学符号（简称数符，又称数码）和一个小数点“.”符号所组成。十进制的计数规则是由低位向高位“逢十进一”。在十个数中没有 10，因为 10 是由 1 和 0 这两个数符组成的，因此，最大数符是 9 而不是 10。10 称为十进制的基数，并且 $10^0=1, 10^1=10, 10^2=100$ 等等。对于任意一个数，在十进制下可以写成以 10 为基数的多项式形式，如 $(3\ 769.57)_{10}=3\times1\ 000+7\times100+6\times10+9\times1+5\times0.1+7\times0.01=3\times10^3+7\times10^2+6\times10^1+9\times10^0+5\times10^{-1}+7\times10^{-2}$。我们称之为十进制数的位置记数法，并称此形式为十进制的多项式表示法。由此可知，十进位的数值实质上是数码与基数 10 的整数次幂之积的和，数3 769.57实即上式的缩写。据考证，十进制跟我们的一双手有十个指头有关。在不发达的原始社会里，双手便是人们随身携带的最方便的计算工具，至今我们仍能够看到在幼儿身上是怎样使用手指作为计数器的。

随着社会的进步，人们的日常工作和生产劳动中的计数并不都是采用十进制的。例如对于时间，1 小时等于 60 分钟，1 分钟等于 60 秒，是六十进制，此外还有八进制、十六进制等等。在自然界和人类生活实践中，还存在着大量

---

* 本文原载于《自动化博览》1992 年第 1 期。

的两种对立状态的现象。例如，开关的“通”与“断”，电位的“高”和“低”，电脉冲的“有”和“无”等等，举不胜举。对于这两种相反的状态，人们选用了两个数符“0”和“1”来表示：若规定“0”表示其中的一种状态，“1”就表示与之对立的另一种状态。例如，“0”表示开关的“通”这一状态，则“1”就表示开关“断”的状态；反过来，用“1”表示“通”，则“0”就表示“断”，这当然也是可以的。

任何一个逻辑电路，包括计算机电路在内，无论它有多么复杂，其输入输出信号都是由一连串的高高低低的电平组成的。这里的电平的高或低是由电子管或晶体管处于开关状态形成的，也就是由“0”和“1”两种状态的变化组成的。

“0”和“1”是两个数符，可以用它们组成一种计数制，称为二进制。二进制只有两个数符“0”和“1”。它的基数是2，计算规则是：每个二进制数自右向左为由低位到高位的计算方向，“逢二进一”。“0”表示没有，加“1”后为1，即$0+1=1$，这和十进位制是一样的。再加“1”呢？因为它是逢二进一，$1+1$要进位得“10”，但这里的“10”已不是十进制中的10了，而是相当于十进制中的2，不过用二进制中的“10”表示十进制的2罢了。再加“1”呢？得“11”，即$10+1=11$，这里的“11”相当于十进制中的3。再加“1”得“100”，它相当于十进制中的4。依次类推。

把二进制中$1+1=10$的进位情况与十进制中的$9+1=10$的进位情况比较一下可知，前者是逢二进一，后者是逢十进一。对于同样由“1111.11”这六个数码所组成的数，在十进制和二进制中各位“1”所表示的值是不相同的：

	1	1	1	1	.	1	1
十进制数：	10^3	10^2	10^1	10^0		10^{-1}	10^{-2}
二进制数：	2^3	2^2	2^1	2^0		2^{-1}	2^{-2}

人们日常生活中，十进制就足够了，但随着科学技术的进步，逻辑电路、计算机电路选用的都是具有两个稳定态的物理元件，所以必须采用二进制，这与人们的习惯是矛盾的。因此人们希望在二进制和十进制之间找出一种全新的计算方法，以部分地解决上述矛盾，使得人们看到这种二进制，最终归结为人们已习惯了的十进制。所以必须学会十进制转换为二进制（或其他进位制），或者是反过来，二进制（或其他进位制）转换为十进制的方法。例如把二进制数1101（记为$(1101)_2$）化成十进制数：$(1101)_2=1\times2^3+1\times2^2+0\times2^1+1\times2^0=2^3+2^2+0+2^0=8+4+1=(13)_{10}$。以下表说明十进制与二进制的对应关系以及二进制基的表示形式：

十进制数	二进制数	二进制基的表示形式
0	0	0×2^0
1	1	1×2^0
2	10	$1\times2^1+0\times2^0$
3	11	$1\times2^1+1\times2^0$
4	100	$1\times2^2+0\times2^1+0\times2^0$
5	101	$1\times2^2+0\times2^1+1\times2^0$
6	110	$1\times2^2+1\times2^1+0\times2^0$
7	111	$1\times2^2+1\times2^1+1\times2^0$
8	1000	$1\times2^3+0\times2^2+0\times2^1+0\times2^0$
9	1001	$1\times2^3+0\times2^2+0\times2^1+1\times2^0$

计算机中的数现在还都是采用二进制进行运算的，通过二进制数与十进制数之间的转换，我们就能够把通常的十进制运算交给机器（计算机）用二进制来进行运算，最后把得到的二进制计算结果转换成十进制数输出。

为什么计算机不采用人们习惯的十进制而要采用二进制呢？从计算机的设计与制造观点来看，二进制比十进制具有某些实用的优点。

第一，二进制容易实现。二进制只用 0 和 1 这两个不同的数码就可以表示一切数，二进制的运算实际上就是关于 0 和 1 的运算。在计算机里，0 和 1 是通过两种对立的稳定状态来实现的。显然，在自然界这种物理元件是普遍存在的，也就是说容易实现的。例如指示灯的“亮”和“暗”是灯的两种对立的稳定状态，如果用一只灯表示一位二进制数，指示灯的“亮”表示“1”，“暗”表示“0”。在计算机中广泛采用电位的“高”和“低”，电脉冲的“有”和“无”来分别表示“1”和“0”。具体实现这些信号的物理元件常用的有开关、穿孔纸带（或穿孔卡片）、磁元件、晶体管或集成电路所构成的单元电路等。总之，对于一位二进制数，在计算机内只要一个物理元件具有两种对立的状态，那么制造具有两个不同状态的元件就比制造十个（或多个）稳定状态的元件容易得多。如果采用十进位制，就有十个数符，而计算机不可能像人一样，一眼就识别这十个符号。因此，在计算机内只能用物理元件的不同稳定状态来表示这些不同的符号，对于一位十进制数就需要一种物理元件具有十种不同的稳定状态，这是很难实现的，因为具有十种不同稳定状态的物理元件不多，而且制造工艺复杂，这在技术上也是困难

的，此外，造价也昂贵。

第二，二进制的运算简单。因为二进制数只有0,1两个数符，它的加、减、乘、除等算术运算规则要比十进制简单得多。譬如二进制的加法规则是：0+0=1，0+1=1，1+0=1，1+1=10，利用此规则可以实现二进制数的加法运算；二进制的减法规则是：0－0=0，1－0=1，1－1=0，10－1=1，利用此规则可以实现二进制数的减法运算；二进制的乘法规则是：0×0=0，0×1=0，1×0=0，1×1=1。可见，乘法规则特别简单，除了1×1=1外，其他情况都等于0，无需“九九口诀表”之类的东西。二进制数的乘法实质上是由“加”（加被乘数）和“移位”两种操作实现的；二进制的除法运算与十进制一样，同样是乘法的逆运算，实质上是由“减”（减除数）和“移位”两种操作实现的。

因此，二进制数的加、减、乘、除等算术运算可以归结为加、减、移位三种操作。在计算机中为了简化设备，往往只设置加法器，而没有减法器。这时，减法是转化为“加负数”的办法来实现的，也就是“加补码”的办法。引进补码是很重要的，因为这样一来，使用加法器就可完成减法运算了。什么叫“补码”？例如在十进制中，要减去9或997之类的数时，通常是采用先减去10或1 000，然后再加上1或3，这样算起来既简便又快捷，这里的补码就是某数与10或1 000之差。例如，因为1=10－9，所以9的补码是1，3=1 000－997，所以997的补码是3。由此可得出由10，100，1 000，…减去某数，其差便是某数的补码。二进制数的补码与十进制数的补码其求法是一致的。简言之，以二进制表示的某数，加上其补码（也是以二进制数表示）后，其和也是10，100，…（均是二进制数）。当引入补码后，二进制数的四则运算就可以归结为加和移位两种操作。二进制数的运算简单这一特点，使利用电子线路来实现自动计算时，计算机的结构可以做得比较简约，也有利于工作可靠性的提高。

第三，采用二进制可以节省存储设备。在计算机中，数是用一些元器件的稳定态来表示的。例如要表示0到999这1 000个数码，采用十进制时需要3位数，每位数有0～9十个设备状态，因此3×10=30；如采用二进制来表示同样1 000个数，则需要10位二进制数，每位数有0或1两种状态，因此10×2=20。所以用二进制数表示时所需设备量要比十进制表示时要少。

第四，采用二进制可以使计算机进行逻辑判断。如上所述，在计算机中，“1”和“0”是由两种对立的稳定状态来表示的，如同“是”和“非”一样，“1”对应“是”，“0”对应“非”，而对应是和非的判断实际是一种逻辑判断。事实上，只要将二进制数运算法则稍作修改，就可以进行二值逻辑运算：逻辑加、逻辑乘、逻辑非。以这些运算为基础，计算机就能够进行逻辑判断和推理，机器也就具有了人类思维的某些功能。

由于二进制具有上述特点，所以计算机采用二进制有着广泛的物质基础，从而获得广泛的应用。发明二进制实在是一件了不起的大事。

采用二进制表示数，是现代计算机的一个重要特点，不过二进制的思想早已有之。据考证，二进制最早出现在中国古代《易经》一书的八卦图中。中国的八卦是用阴、阳两种概念来试图说明一些自然现象的产物，它用符号“--”代表阴，用符号“—”代表阳，用这两种符号的不同排列，来表示天、地、风、雷、水、火、山、泽等八种自然现象。这阴阳两种一负一正的概念正好与二进制数码相对应。《易经》中的符号系统实际上就是二进制的符号系统。17 世纪德国数学家莱布尼茨在设计他的计算机的时候，就想出了二进制的想法，他的思想即来源于中国的八卦，可惜他的愿望未得以实现就离开了人间。19 世纪，英国数学家布尔(George Boole)创立了当今有名的布尔代数，然而直到 20 世纪，人们在控制系统中，在计算机的应用研究中才注意到布尔理论的作用，布尔代数成为计算机理论的基石。

二进制真是既古老又年轻。□

3 推理机制与推理方法*

推理是人们在长期社会实践中，对客观事物的联系与关系的一种反映，可以从已有的知识推导出新的知识。所以，推理也是人类特有的智能活动，是逻辑演绎的过程。

一般说来，推理包括两种判断，一种是已知的判断，称为前提；另一种是由已知的判断推出的新判断，称之为结论。

一　推理的机制与方法

根据人脑思维方式上的特点，可以把推理分为演绎推理、归纳推理、类比推理等。演绎推理是一种充分置信推理，而归纳推理则不同，它是主观上不充分置信推理，当然也有一种可认为是部分地转移置信度的情况。从推理过程的方向上，即前提与结论之间的路径关系上，又可分为正向推理、反向推理、混合推理三种。

1. 演绎推理。演绎是从一般的公理定律出发，采用形式逻辑的方法来认识事物的一种脑功能。它包括数值计算和逻辑推理。

演绎推理是从已知的判断出发，通过演绎推出结论的一种推理方式，其结论就蕴含在已知的判断中，所以演绎推理是一种由一般到个别的推理。如果前提百分之一百地可信，则通过演绎推理得出的结论也百分之一百地可信。从此意义上说，演绎推理可认为是一种“保真变换”。

*本文原载于《自动化博览》1997年第6期。

一般地，如果前提的置信度为 p，则通过演绎推理得出的结论也应该具有置信度 p。

三段论法是演绎推理的核心。三段论法是由两个前提（大前提、小前提）和一个结论组成的。它包括直言三段论法、假言三段论法、选言三段论法和二难推理等类型。在目前知识系统中大量使用的是假言三段论法，又称假言推理。如果使用 if—then 的形式，则假言三段论法的基本形式有肯定式和否定式两大类。

肯定式：若知"如果 A 则 B"且 A 为真，则可推出 B 为真。

否定式：若知"如果 A 则 B"为真和 B 为假，则可推出 A 为假。

演绎推理还包括谓词演算等形式推理，如定理机器证明中的归结原理。

归结原理是 1956 年美国数学家鲁滨逊（Robinson）根据 1930 年法国的艾尔勃朗（Herbrand）定理而提出的。由于这种推理方法克服了一阶逻辑的最大难点，解决了由量词而引起的麻烦，容易改进、完善，便于用程序在计算机上实现，所以，自 1956 年以来，是一直被采用的方法。它的基本内容是：若命题 P 成立，则命题 Q 成立；若命题 P 不成立，则命题 R 成立。考核这两个命题是否为矛盾时，可以合并为考核新命题 Q 和 R 是否为矛盾，由此扩展下去，直至推出矛盾，即原命题的否定不成立，从而证明原命题成立。简要地说，这种方法是把应该证明的命题加以否定，然后导出矛盾，通过证明其否定为假来证明原命题为真。这个推理就叫归结。从狭义来看，这一原理只涉及形式逻辑中命题证明的一种方法，实际上却可以用到问题求解的任何场合。

归结原理的重大意义在于，这种证明方法标志着人工智能的研究从机器仿人的求解目的转向研究机器求解问题的方法，即完全摆脱了过去"仿人"的证明方法的束缚，从而使证明方法更适合于计算机的特点，使定理证明的机械化变为现实。

2. 归纳推理。归纳是从部分到整体，从特殊到一般，从个别到普遍的一类推论的行为、过程或结果。归纳作为一种思维方法，是逻辑学的组成部分，也是哲学研究的对象之一。归纳作为一种思想方法来研究，形成了归纳逻辑，归纳推理就是它的主要研究对象。

所谓归纳推理，就是从足够多的事例中归纳出一般性知识的推理，是从个别到一般的推理。归纳推理一般包含由特殊到一般的归纳、由特殊到特殊的归纳和统计三段论法。

归纳推理与演绎推理恰好相反，即演绎推理是从前提到结论，而归纳推理则是从结论到前提，是从给定的事实出发，推论出具有一定可信度的一般断言或假说。它可以说明给定的事实并可推出新的事实。如果说演绎推理是"正向

推理”，则归纳推理就可以说是“逆向推理”。

归纳推理还有一个很重要的区别于演绎推理的特点是：由归纳推理得出的结论是否确实，是无法判断的，通常只能从一定的置信度予以接受。因此归纳推理是属于所谓主观的不充分置信的推理一类。即“它能从一个具有一定置信度的前提推出一个比前提置信度低的结论”。归纳推理除穷举归纳和数学归纳外，一般归纳推理只是“保假”的。即若归纳依据的前提错误，则结论错误；但前提正确时结论也不一定正确，应按某一标准选取最好的作为归纳结果。

归纳推理是人类智能行为的重要表现，人们在学习过程中常常自觉或不自觉的运用着归纳法。由于归纳推理增加知识，因此也称归纳推理为归纳学习。尤其是在进行机器学习的研究中，基本归纳的学习是一种很有效的学习方法。除了纯机械式的学习外，也都需要借助于归纳推理规则。有的学习过程本身就是一种归纳推理过程。

3. 类比推理。类比推理有时也称为类比学习。类比是人类思维的重要方式之一。所谓类比推理，就是在两个或两类事物有许多属性都相似或相同的基础上，推论出在其他属性上也相似或相同的一种推理。

类比推理有多种形式，经典的类比推理是指：设事物 a 有 n 个属性 $p_1,p_2,\cdots,p_n$，事物 b 也有这 n 个属性，并且已知 a 还有第 $n+1$ 个属性 p_{n+1}，则可类比出 b 也有第 $n+1$ 个属性 p_{n+1}。这就是用类比推理推出的新知识。

由于类比推理可以从事例中推出知识，因此它通常被用作知识系统的知识获取工具。许多科学发明和发现就是通过类比推理来实现的。例如在化学领域里，有不少化学元素的发现就是利用了类比推理。

二 推理的路径与方法

推理是运用各种推理规则进行思维的过程，是求解问题的过程。问题求解的质量不仅依赖于所采用的求解方法，也依赖于求解问题的策略，即推理的控制策略。控制策略包括推理方向、推理路线、推理的效果与效率等。按推理进行的路线与方向，推理可分正向推理、反向推理和混合推理。

1. 正向推理。正向推理是从最初的原始条件（或原因）出发向着所要解决的问题的最终目标（结论）前进的一种推理方式，它是用初始条件来驱动推理过程的。在这种推理过程中，通过对输入的条件进行处理（包括模式匹配、比较等）而得到中间结果，再对这种中间结果进行处理（模式匹配、比较）。反复进行这种进程，直至得出最终目的（结论）来。由于它是从数据到结论的过程，所以

这种推理策略也称之为数据驱动推理策略或数据驱动方式的推理。

简而言之，正向推理是以前提（或原因）推结论（或结果）的推理。由一步正向推理所得出的结论又可能成为另一规则的前提，从而又可接着推出另一新结论。如此“头—尾”相连进行推理，可以得一条所谓“正向推理链”。

正向推理的优点是比较直观，相对来说比较容易实现，缺点是无明确的目标，开销大。

2. 反向推理。反向推理的方式正好与正向推理的方式相反，即以目标来推动推理过程的一种模式。这是在推理过程中首先从目标出发，不断地找出满足目标的所有情况、条件，反向地向着最初的情况、条件逼近，重复此项工作，直至达到或符合最初的（或最原始的）情况条件。因为它是由结论（目标）到数据的，所以又称作目标驱动推理策略。

简言之，即从结论（或结果）反推可能的前提（或原因），这种推理方式称为反向推理。同样，反向推理过程也会得到一条“反向推理链”。

反向推理目标明确，但实现起来比较困难，尽管它的开销也许会少些。

3. 混合推理。把正向推理和反向推理这两种推理结合起来形成一种推理，就叫混合推理。它从初始前提（或原因）和最后结论（或结果）两头分别采用正向与反向推理，使形成的正向推理链与反向推理链在中间某处会合。

还有一些其他形式的推理，如单调推理、非单调推理、不精确推理等，使用时要根据问题的实际情况，选择一种合适的推理。□

4 美妙的黄金分割*

数学家对最优化问题的研究，可以说源远流长。公元前 500 年，古希腊的毕达哥拉斯就发现了所谓“黄金矩形”，即矩形的长与宽的最佳比例为 1.618∶1（或 1∶0.618）。这一比例被称作黄金分割比。$\frac{\sqrt{5}-1}{2}$（≈0.618）被称为黄金数。在数学上，它可作为一种有效的一维搜索最优化方法，简称0.618法。在平面几何中，它又被称作“中末比”，即在已知线段 AB 上取某点 C，使$\frac{AC}{AB}$之数值为黄金数。这是一种具有特殊性质的比值。在代数上，$\frac{\sqrt{5}-1}{2}$恰好是方程 $x^2+x-1=0$ 的正数根，而$\frac{\sqrt{5}+1}{2}$（≈1.618）则是 $x^2-x-1=0$ 的正数根。

据说，毕达哥拉斯某日路过一个铁匠作坊，听到十分悦耳的丁丁当当的声音，于是他便测量了铁锤与铁砧的尺寸，并发现它们之间有一定的比例。为此，他把一条直线段分为两段，经过反复比较，最后终于确认 1∶0.618 的比例最优雅。

人们在探索自然美和艺术美的过程中，也发现黄金分割比具有一种非常悦目的美。古希腊和欧洲的建筑师及美术家认为，在建筑、人像雕塑和绘画中，应用黄金分割比将使建筑与艺术最为优美和协调。例如，在建筑物造型方

*本文原载于《百科知识》1989 年第 2 期。

面，人们在高塔的黄金分割点处建造楼阁或设置平台，便能使平直、单调的塔身变得多姿多彩，而在摩天大楼的黄金分割点布置腰线或装饰物，则可使整座楼显得宏伟而雅致。在雕塑作品中，艺术家认为人的形体以下肢与身高之比为0.618时最美。古希腊时期的塑像如智慧女神雅典娜、太阳神阿波罗等，都是采用这种身段比例。在音乐会上，报幕员如果站在舞台正中，则会显得呆板，但若站在舞台长度的黄金分割点的位置上，观众在台下就会感觉很匀称，而且声音的传播效果也最好。在演奏乐器时，如果将二胡上的“千斤”放在琴弦的黄金分割处，则音色可达到最佳。鉴赏家们还认为，用“中末比”来规划矩形的两边，形状十分优美。因此，日常生活中的许多艺术构思，常常需考虑到黄金分割的特点，特别是一些矩形物品，如书本、柜橱、门窗等，应用黄金分割比可给人以和谐美观之感。

饶有趣味的是，植物学家们观察到某些植物的生长也是按黄金分割的序列排列的。当一株嫩芽抽枝吐叶时，如果从这株嫩枝的顶端看下去，可以看到叶子的排列成一对数螺线，而叶子在螺旋线上的距离恰好符合黄金分割。人们按照车前草叶形排列的数学模式设计出来的螺旋形高楼大厦，每一个房间都能得到充足的阳光。无独有偶，向日葵的种子也是按特定的螺线弧排列的，它们在螺旋线上的距离竟也服从黄金分割规律。

此外，人们还发现一个底角为 72°的等腰三角形，其底边与腰的比值即为“中末比”，而正五边形和正十边形恰与这种三角形有关，进而可推出正十边形的边长与半径之比符合“中末比”。利用这一关系，便可实现正多边形的作图。

13 世纪，意大利数学家斐波那契在其著作《算盘书》的修订本中提出一个有趣的“兔子问题”，为黄金分割比大增光彩。这个问题的大意是：已知一对兔子生下后第二个月就开始生小兔子，那么，从刚出生的一对兔子开始算起，满年时可以繁殖成多少对兔子呢？

由第一个月到第十二个月兔子的对数分别为

$$1,1,2,3,5,8,13,21,34,55,89,144。$$

答案是满一年时间可以繁殖 144 对兔子。

上面这个数列称为斐波那契数列。它有这样一个特点，任一数都是前面两个数的和。

若一数列 $F_1, F_2, \cdots$ 满足任一数都是前两个数的和，即

$$F_n = F_{n-1} + F_{n-2}, n \geqslant 3,$$

则称此数列为斐波那契数列，其中每一项的数称为斐波那契数。其通项公式为

$$F_n = \frac{\sqrt{5}}{5}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]。$$

上式右边第二个小括号内即为“中末比”的负值。最后，可得出结论：斐波

那契数列中前后两项的比值$\frac{F_n}{F_{n+1}}$是以黄金数$\frac{\sqrt{5}-1}{2}$为极限值。

这个结果表明，n 值越大，则数列中前后两项的比值就越接近$\frac{\sqrt{5}-1}{2}$。如 1∶1＝1，1∶2＝0.5，2∶3≈0.667，3∶5＝0.6，5∶8＝0.625，8∶13≈0.615，13∶21≈0.619，…，34∶55≈0.618，…

一些植物学家认为，自然界中花朵的花瓣数目大多为 3，5，8，13，21，34，…不过，造成这种结果的原因何在，目前尚未查清。

体现“中末比”性质的斐波那契数列，在数学界风靡了几百年，后来人们由此提出了一种具有实用价值的“优选法”，并立即在工农业生产中发挥出了巨大的作用。通常，人们试制一种新产品，或者选取合适的配方、配比，或者改革一项旧工艺，总要先做科学实验。黄金分割法就是一种以较少的实验次数尽快找到最佳生产方案的科学实验方法。过去，人们曾经花费几年甚至几十年的时间，做几百次乃至上千次的实验，才能获得某种科学成果。黄金分割法问世之后，仅用几天甚至几个小时就可获得满意的结果。因此，黄金分割法又叫做优选法。20 世纪 70 年代，我国著名数学家华罗庚在全国许多大中城市的工厂向群众宣讲和推广优选法(即 0.618 法)，使广大工程技术人员获得了数以百计的优选成果，在国家不增加投资、设备和人力的条件下，收到了巨大的经济效益。

黄金分割普遍被看做是最佳的比例。数学家认为它简洁，美学家认为它和谐，艺术家认为它妙不可言。人们可以发现，物质世界的组成、大地万物的诞生以及世间的许多事物，都和黄金分割有着千丝万缕的关系，似乎它不仅是哲学的领悟、数学的技巧和艺术的完美之间最惊人的结合，而且还是构成世界与宇宙原动力的内部规律。它体现着人类能感觉到的蕴藏在这个世界之后的神奇结构和深奥理性。黄金分割真是贵如黄金！□

模糊数学：一门研究模糊性现象的学问

1 模糊数学及其应用*

一　清晰与模糊

为了识别事物，人们总要依据一定的标准对它们进行分类。在现实世界中，人们可以依据精确的标准把许多事物分为彼此界限分明的类别，每个事物要么属于某一类，要么不属于该类，非此即彼，明确肯定，绝对不模棱两可。我们把事物有明确类属的这种特性，称为清晰性，把这类事物称为清晰事物。例如人是动物、地球是行星等，都是清晰事物。

但是，对于另外一些事物，我们却无法找到精确的分类标准，对于某一事物究竟属于哪一类很难作出明确肯定的断言。高山、大河，远亲、近邻等就是这种事物类。由于这种事物从属于某一类到不属于该类是逐步过渡而非突然改变的，不同类别之间不存在截然分明的界限，因而不同人对同一事物可能作出不同的归类。我们把事物的这种类属的不清晰性称为模糊性，把这类事物称为模糊事物。

另外，从客观事物本身来说，它们也不是孤立的、僵死的，而是相互联系、不断转化的，所以事物之间的界限既是分明的，又是不分明的，绝对分明和固定不变的界限是不符合客观事物的本性的。对此，恩格斯做了很精彩的说

* 本文完成于 1996 年。

明。他说:"辩证法不知道什么绝对分明和固定不变的界限,不知道什么无条件的普遍有效的'非此即彼'。它使固定的形而上学的差异互相过渡,除了'非此即彼',又在适当的地方承认'亦此亦彼',并且使对立互为中介。辩证法是唯一的、最高度的适合于自然界的这一发展阶段的思维方法。"再从人类认识的本质来看,人的认识基本上是模糊认识,是有其客观原因的,而且是不可避免的。人类认识的这种模糊性不但无害,在许多情况下反而有益。一方面,人们对客观事物的这种模糊认识,在一般情况下,对于所要完成的特定任务来说是足够精确的了。另一方面,正是这种模糊认识,使人能够及时地灵巧地处理各种复杂问题。模糊语言、模糊概念有时在解决实际问题中比非模糊语言、非模糊概念能发挥更大的作用。建立在二值逻辑基础上的电子计算机,对确定的问题用确定的规则进行演绎的能力是很强的,但他在模糊现象面前就无能为力了。因此非常需要将人的认识的这种模糊性赋予机器,以提高电子计算机的智能水平,就是亟待我们解决的问题了。

模糊数学就是用数学方法研究和处理模糊性现象的一门新兴学科。

二　由普通集合到模糊集合

在数学上,一个清晰事物类可以表现为一个集合。集合是现代数学的一个重要基础概念。在传统或经典的集合论(即康托尔集合)的元素与集合的关系,一个元素属于或不属于一个给定的集合,二者必居也只居其一。设一个集合 A,元素 x,按照传统的集合理论,属于或不属于在逻辑上看做是一种值。若 x 属于 A(记为 $x\in A$)用值"1"表示,x 不属于 A(记作 $x\notin A$)用值"0"表示,这样,集合可以用一个特征函数表示。集合 A 的特征函数记为 $f_A(x)$,定义为

$$f_A(x)=\begin{cases}1,当\ x\in A;\\0,当\ x\notin A。\end{cases}$$

这种关系属于二值的、绝对化的规定,是对清晰事物类属关系的科学抽象,刻画了清晰事物要么具有某种性态,要么不具有该种性态的"非此即彼"的特征。一个清晰事物类就是一个普通集合,以普通集合作为描述清晰事物的数学模型,有关的数量关系即可得到精确的描述。但也正是这一条假定,使普通集合本质上不能描述类属不分明的模糊事物。要克服这个困难,就需要推广集合概念。

为了要给模糊事物以适当的集合描述,1965 年,美国控制论学家、加利福尼亚大学教授扎德(Zadeh)把经典集合中,元素对集合要么"属于"要么"不属于"的分明关系推广到元素按"一定程度"隶属于某个集合的关系,这个"一定程度"

叫做元素对集合的隶属度。以“～”作为模糊化记号，模糊集合记为$\underset{\sim}{A}$，$\underset{\sim}{B}$，$\underset{\sim}{C}$等。集合U表示论域，即所讨论的问题范围，它是一个普通集合。

设$\underset{\sim}{A}$是论域U上的模糊集合(要注意模糊集合$\underset{\sim}{A}$是以集合U作为参考集的，所以严格说来，模糊集合$\underset{\sim}{A}$应称为模糊子集)。U中百分之百地属于$\underset{\sim}{A}$的元素对$\underset{\sim}{A}$的隶属度为$\mu=1$，百分之百不属于$\underset{\sim}{A}$的元素对$\underset{\sim}{A}$的隶属度为$\mu=0$，其余的元素对$\underset{\sim}{A}$的隶属度用介于0和1之间的实数μ来表示，较大的μ值表示较高的隶属度。那么，模糊集合就得到一种定量的刻画。

定义　论域U上的模糊集合$\underset{\sim}{A}$是用一个从U到实区间[0,1]的函数$\mu_{\underset{\sim}{A}}(x)$来刻画的，$\mu_{\underset{\sim}{A}}(x)$叫做模糊集合$\underset{\sim}{A}$的隶属函数，函数值$\mu_{\underset{\sim}{A}}(x)$代表元素$x$对$\underset{\sim}{A}$的隶属度。

为了方便，有时将模糊集合$\underset{\sim}{A}$的隶属函数记为$\underset{\sim}{A}(x)$。

这样，普通集合用特征函数来刻画，模糊集合用隶属函数来刻画，特征函数的值域为集合{0,1}，隶属函数的值域为[0,1]。隶属函数是特征函数的推广，它是模糊集合乃至整个模糊数学的最基本的概念之一。

模糊集合把普通集合中元素隶属于集合与否的“非此即彼”的绝对性，扩展为元素对集合隶属程度的“亦此亦彼”性。

隶属函数也称为资格函数或从属函数，刻画的是元素从属于集合到不属于集合的渐变过程，亦即隶属度在论域上的分布。所谓模糊性，就是元素对集合属于关系的不分明性，属于程度的连续过渡性。

隶属度是模糊集合论应用于实际问题的基石。一个具体的模糊性对象，首先要写出它的切合实际的隶属函数，才能应用模糊数学方法作具体的定量分析。因此确定隶属函数是十分重要的。但隶属函数是人们主观构造的，是一个构造函数，它包含着人们的主观成分及技巧，人们可以根据不同情况所提供的部分信息而构造这样或那样的一个模糊集合的隶属函数，不同的构造代表不同的识别。正确构造隶属函数是应用模糊数学方法的关键，但这个问题至今尚未获得令人满意的解决。应用于实际问题时，可根据具体情况采用适当方法来确定隶属函数。一种较有效的方法是通过模糊统计试验来确定隶属函数，通常的理论文献的例子，大多用推理方法近似指定隶属函数。

模糊集合的表示法　给定有限论域$U=\{x_1,x_2,\cdots,x_n\}$，$\underset{\sim}{A}$为U上的模糊集合，μ_i为x_i对$\underset{\sim}{A}$的隶属度，通常按扎德的记法，将$\underset{\sim}{A}$表示为

$$\underset{\sim}{A}=\frac{\mu_1}{x_1}+\frac{\mu_2}{x_2}+\cdots+\frac{\mu_n}{x_n}。$$

上式右边并非算术中分式相加。$\frac{\mu_i}{x_i}$中的分母是论域U中的元素，分子是该

元素对 $\underset{\sim}{A}$ 的隶属度，取分式的形式只表示 μ_i 与 x_i 的对应关系。"+"号表示将各项汇总，表现集合概念。若 $\mu_i=0$，在上式中可以略去该项。

三　模糊集合的运算

模糊集合的运算可以转换为相应的隶属函数的运算。如同普通集合的运算，模糊集合最基本的运算也是并、交、补三种，运算符号也采用∪、∩、∁。设 $\underset{\sim}{A}$,$\underset{\sim}{B}$ 为论域 U 上的模糊集合，$\underset{\sim}{A}$ 和 $\underset{\sim}{B}$ 的并集记作 $\underset{\sim}{A}\cup\underset{\sim}{B}$，交集记作 $\underset{\sim}{A}\cap\underset{\sim}{B}$，$\underset{\sim}{A}$ 的补集记为 $\underset{\sim}{A}^{\complement}$。

人脑在分析处理模糊事物时，经常进行求并、交、补的运算。例如，中到大雨＝中雨∪大雨，瘦高个子＝瘦个子∩高个子，不满意＝(满意)$^{\complement}$ 等。

设 $\underset{\sim}{A}$,$\underset{\sim}{B}$ 为论域 U 上的模糊集合，以符号∨表示取最大值，∧表示取最小值，定义：

并　$\underset{\sim}{A}\cup\underset{\sim}{B}=\max\{\underset{\sim}{A}(x),\underset{\sim}{B}(x)\}=\vee(\underset{\sim}{A}(x),\underset{\sim}{B}(x))=\underset{\sim}{A}(x)\vee\underset{\sim}{B}(x),\ \forall x\in U$；

交　$\underset{\sim}{A}\cap\underset{\sim}{B}=\min\{\underset{\sim}{A}(x),\underset{\sim}{B}(x)\}=\wedge(\underset{\sim}{A}(x),\underset{\sim}{B}(x))=\underset{\sim}{A}(x)\wedge\underset{\sim}{B}(x),\ \forall x\in U$；

补　$\underset{\sim}{A}^{\complement}=1-\underset{\sim}{A}(x),\ \forall x\in U$。

上述模糊集合的运算都有现实原型。通常所谓两利相权取其大，两害相权取其小，就是取最大值和取最小值运算的一种原型。

四　模糊关系

在日常生活中，"关系"是最普遍的名词，用来形容人与人之间、各种不同事物之间的联系，如"父子"关系、"朋友"关系等。在数学上所研究的是一种抽象的概念，它是用严密的数学语言来定义的。设 Z,Y 是两个实数集合，在 Z 中取一个元素 x，又在 Y 中取一个元素 y，按 x 在前、y 在后的次序搭配起来所成的有序对(记为 $<x,y>$)的全体所构成的集合称为 Z,Y 的直积，记为 $Z\times Y$，即 $Z\times Y=\{<x,y>\mid x\in Z,y\in Y\}$。直积 $Z\times Y$ 的任何一个子集 R 称为 Z,Y 之间的二项关系，简称 Z 到 Y 的关系。当 $Z=Y$ 时，$Z\times Z$ 的子集 R 称为 Z 中的二项关系，简称 Z 中的关系。

除了可用普通集合描述的那种清晰确定的关系之外，客观事物之间还存在着大量不大清晰或不完全确定的关系。侦察员常依据罪犯留下的脚印大小推测罪犯的身高，因为人的身高与脚长之间有某种近似的关系。人的面貌相似也是这种关系。语言学中的词义相近，数学中的"远大于"、"充分接近"等都是这类不清晰的关系。这种事实上存在但又不那么清晰确定、不能用普通有序对集合描述的关系，我们称之为模糊关系。模糊数学中的模糊关系，是上面通常关系的推广。

扎德给这种广泛存在的模糊关系(原型)提供了一种集合描述。一般地，有如下定义：

所谓从集合 U 到集合 V 的模糊关系 $\underset{\sim}{R}$，是指直积 $U\times V$ 上的一个子集合 $\underset{\sim}{R}$，由隶属函数 $\mu_{\underset{\sim}{R}}(x,y)$ 刻画，函数值 $\mu_{\underset{\sim}{R}}(x,y)$ 代表有序对 $<x,y>$ 具有关系 $\underset{\sim}{R}$ 的程度。

这是二元模糊关系的数学定义，多元模糊关系亦可类似地定义。

有限论域上的模糊关系都可以用模糊矩阵表示，每个模糊矩阵都代表一定的模糊关系，布尔矩阵是模糊矩阵的特殊情况，代表普通关系。

因为模糊关系是模糊集合，因此可以对它定义相应的并、交、补等运算，所以自然可以根据模糊关系中的概念和运算，确定模糊矩阵的相应概念和运算，如矩阵的并、交、补等运算。

五　模糊逻辑

模糊逻辑原则上是一种模拟人的思维过程的逻辑，是对二值逻辑的扩充。模糊逻辑要考虑被讨论对象属于某一类的程度。一个命题可能亦此亦彼，存在着部分真和部分伪，其最大特点就是用它可以比较自然地处理人的概念。

作为传统逻辑的推广，模糊逻辑是一种非标准逻辑，目前对其尚无统一的确切定义。在更广泛的意义上理解，模糊逻辑是模糊数学进行推理的逻辑工具，它把传统数学从二值(0 与 1)逻辑扩展到真值取闭区间[0,1]上连续值的无限多值逻辑(连续值逻辑)上来。在模糊逻辑中，以 x,y 等表示简单命题的真值，经过逻辑运算后，可算出复合命题的真值，这里模糊变量 $x,y\in[0,1]$，模糊变量之间的基本运算作为二值逻辑运算的扩展，定义为：① 逻辑并(析取)：$x\vee y=\max\{x,y\}$；② 逻辑交(合取)：$x\wedge y=\min\{x,y\}$；③ 逻辑补(否定)：$x^{c}=1-x$。

六 模糊数学应用简介

模糊数学的产生把数学的应用范围从精确现象扩展到了模糊现象的领域。20世纪70年代以来，模糊集合的概念逐渐被人们接受，模糊数学的研究得到迅速的发展。目前已深入到拓扑、代数、数理逻辑、概率论、控制论、信息论、决策论、对策论、最优化等领域，出现了模糊拓扑、模糊群论、模糊逻辑、模糊概率、模糊控制、模糊信息、模糊决策、模糊对策、模糊优化等新分支。理论正在不断地完善，应用范围也日益广泛。20世纪80年代至90年代中期，世界许多国家投入了大量人力物力，模糊控制作为一项新的工程技术在日本、德国、美国、中国等得到广泛的应用。“模糊产品”已经涌入到社会生产和人们的日常生活中，国际上的“模糊产品热”正在兴起。如日本日立公司使用模糊逻辑计算机控制地铁系统，使得加速和刹车变得非常平稳。20世纪80年代以来，日本模糊控制理论及技术与高新技术相结合已经转化为强大的生产力。模糊技术在日本得到了成功的应用，所掀起的“模糊技术热”方兴未艾，模糊控制已使日本自动化迈上了一个新台阶。在日本高度工业化的基础上，加上先进的模糊技术，创造了巨大的经济效益和社会效益。1988—1991年，日本采用模糊技术的新一代家用电器产品上市，是模糊技术工业应用的显著成果。这些家电产品有模糊控制的全自动洗衣机、电冰箱、电烤箱、空调、摄录一体化摄像机、自动电话、自动热水器等等。1987－1993年，日本将模糊控制理论及技术应用于工业生产过程中，使石油工业、冶金、水力电气等工业企业，在节能降耗、提高生产率和产品质量方面取得了显著成绩。在日本，模糊技术的应用已经带来了巨额利润，单是模糊控制芯片一项，1993年销售额就达6亿美元。模糊技术在金融证券部门的开发和应用尤为盛行。日立公司研制的股价预测系统可根据过去股价的变化取出经常出现的20种股价的特征，进而预测未来股价动向。现在证券公司采用普通计算模式对两周后股价动向进行预测，准确率约为40％，而采用模糊技术的预测系统可达60％－70％。这种模糊股价预测系统不是根据单纯的模糊推论，而是应用了根据数值对股价变化进行判断的理论，因而它成为股票商从事交易的新手段。美国是模糊控制理论专家扎德的故乡，20世纪90年代，美国的单片技术得到了迅速的发展，给模糊控制技术在军事工程方面的应用打下了硬件基础，使模糊控制成为美国军事工程中90年代的热点技术之一，为世人瞩目。据报道，1992年美国已把模糊技术应用到信息工程、图像识别、人工智能、空间飞行、卫星与导弹的控制系统，取得显著成效。

我国在模糊理论及模糊控制方面的研究，享有很高的国际地位，进入了“国际四强”（日、美、加、中）的行列。据人民日报 1995 年 2 月 1 日报道，北方工业大学模糊控制研究所于 20 世纪 70 年代到 80 年代初完成了“模糊控制理论”的研究，并在冶金工业窑炉控制等领域的应用中取得了较好的成果。80 年代以来，他们又经过多年的努力，于 1990 年推出“快速自寻优模糊控制理论及其应用技术”，并将其应用于工业过程控制、汽车安全行驶防撞控制及家电产品的控制等领域，已形成了系列产品，取得了明显的经济效益和社会效益。该控制器用于工业生产控制，可降低能耗 20%－25%，提高产品成品率 15%－20%。他们推出的智能化的自寻优模糊控制器，在响应速度、控制精度和控制功能上，已大大超过国外的普通控制器，达到了国际领先水平。此外，北京师范大学数学系汪培庄教授 1988 年指导他的博士研究生，采用我国自己的数学构思和电路设计，研制成功了模糊推理机，利用它进行倒摆控制试验，可以灵活地控制不同形状的物体，甚至可以控制不规则的柔性物体，如花束、树枝等。这种控制具有高速、实时的特点，并且具有很强的抗干扰性，专家们认为它将在海上钻井稳定控制、火箭姿态控制、飞机安全着陆、化工过程控制等国防技术和民用技术领域开辟新的应用前景。

总之，模糊数学的建立和发展，可以在一定程度上弥补经典数学和统计数学描述客观世界的不足，并为我们探索大脑处理模糊事物的功能，让机器表现出更大的“活性”，提供了有力的数学工具。模糊数学在人工智能及其他科学（如机器人学、医学诊断、心理学、生物学等）中广泛而有效的应用，雄辩地说明人脑的认识确实具有灵活性、模糊性。□

2 模糊逻辑及其发展*

最近，日本三洋公司宣布，它已研制成功新一代利用模糊逻辑元件和其他先进计算机电路的智能摄像机，能拍摄出清晰度更高和图像更稳定的录像带。大多数计算机都是以“是或非”这种形式来处理信息的，对摄像机来说就是“黑或白”、“近或远”，但模糊逻辑元件能使计算机对“程度”进行思考，如“中灰”、“相当远”等。而且模糊逻辑元件能使摄像机更好、更快地对条件变化做出反应，更加准确地自动聚焦和自动曝光。

利用模糊理论研制成功的模糊逻辑计算机一问世便受到人们的青睐。把它安装在地铁系统，使列车的加速和刹车变得非常平稳。日本东芝、日立等公司在电梯上争先应用模糊逻辑，使平均等待时间缩短 10%到 15%。乘客等待时间超过 1 分钟以上的次数减少 30%到 40%。用模糊逻辑元件组装的电视机具有“主动信号校正”的功能，对颜色、亮度和对比度每隔 $\frac{1}{60}$ 秒比对一次。它将画面上的 248 个点同存储在机内的 40 幅“完美”的画面的图像库进行比较，寻找相似而不必是相同的条件。模糊逻辑在人工智能、机器人、专家系统等领域都有广泛的应用前景，它是多值逻辑的发展，又是模糊推理的基础。那么，什么是模糊逻辑呢?

逻辑是研究人的思维形式和规律的科学。19 世纪时，英国数学家布尔和哲学家罗素等人把数学方法应用于逻

---

*本文原载于《自动化博览》1993 年第 2 期。

辑的研究，出现了一门数学与逻辑学相互渗透的学科——数理逻辑。在这门学科中，一个可以分辨真假的陈述句叫做命题。一个命题非真即假，或非假即真，二者必居其一。一个命题的真或假叫做它的真值，由于它仅取真、假二值，故在数理逻辑上也称二值逻辑。通常用“1”代表真，“0”代表假。但1和0仅表示两种相互对立的稳定状态，不具有数学中的那种大小关系。至于其物理意义则随具体问题的不同而异，可以是{真、假}，{开、关}{通、断}等。显然，通常的数学方法不能适用于这类问题，它需要一种特殊的数学工具。在历史上，布尔首先提出了逻辑代数的理论，因此也称为布尔代数。现在的计算机就是建立在二值逻辑的基础之上的。但进入20世纪60年代后，人类研究的领域越来越复杂，如人脑系统、智能系统等。复杂的东西是难以精确化的。虽然计算机的运算速度快得惊人，但处理模糊现象却无能为力。例如看电视，要把图像调得清晰些，这对一个小学生来说都是轻而易举的事，但若让计算机去完成则是一个大难题。可见当计算机的触角伸展到模拟人脑思维过程的领域时，二值逻辑判断的绝对化——非真即假，便成了一个障碍。因为人脑的思维活动（包括分析、综合、推理、判断等）并不是绝对的这种抉择，而是在“真、假”和“是、非”之间存在着过渡状态，处于过渡中的思维具有模糊性。所以为探索人类自然语言（即具有模糊性特点的语言）的形式化表达，由此真正解开逻辑思维之奥秘，使计算机能模拟或代替人的某些智能活动，就需要把二值逻辑推广和发展成多值连续形式的逻辑。这是当代科技领域的前沿课题之一。国内外不少学者现在也都希望逻辑元件能在0与1之间连续取值，并认为这样才更能反映人脑的思维。因此，模糊逻辑是一种形式化的连续值逻辑。在模糊逻辑中，模糊命题的取值不是绝对的“0”或“1”，而是在闭区间[0,1]中的某个实数，其值的大小表征“模糊性”或“真假性”的程度。它不仅为近似推理或似然推理提供了一个基础，而且模糊逻辑在开关理论、信息处理以及复杂工业生产过程中的语言控制等方面也都有广泛的应用。

1965年，美国控制论学家扎德教授发表了题为《模糊集合》的论文，标志着模糊数学的诞生。模糊数学的主要特点，一是能定量地处理影响分析和设计中的种种模糊因素，使分析的结果和设计的方案更符合客观实际和更为优化合理；二是能充分考虑事物的中介过渡性质，浮动地选取阈值，从而能给出一系列不同水平下的分析结果和设计方案，为人们提供可以选择的余地。此外，它也为模糊逻辑的研究提供了有力的数学工具。尽管至今人们对这一学科的一些问题还有分歧意见，但它已渗透到一些数学的分支，并对当前科学技术的研究产生了重大的影响。事实证明，它是处理客观事物的某些模糊性的一种有效的数学工具。今天，对许多人来说，模糊数学或许已不再是神秘而陌生的科学了。

1974 年，英国工程师马丹尼(Mamdani)首先利用模糊控制语言研制成功模糊控制器，并把它应用于蒸汽机和锅炉的控制等方面，使模糊逻辑得到了飞跃的发展。现在，糊糊逻辑的主要内容有模糊命题、模糊逻辑联结、似然推理等。

模糊命题是模糊逻辑的一个最基本的概念，是模糊逻辑的基础。什么叫模糊命题？如陈述句“今天真热”，“真热”的概念是模糊的，因为气温的变化是连续不断的，不存在一个突然的转折，致使“热”与“真热”没有一个截然分明的界线，所以称“今天真热”是模糊陈述句。这类陈述句无确切的真假可言，因此，我们把含有模糊概念的陈述句叫模糊命题，它的取值已不再是绝对的真或假，而是真、假的程度。例如在日常生活中，我们常用“非常真”、“真”、“有点真”、“有点假”、“假”、“完全假”之类的表达方法来表征一个陈述句的真假程度，但这还只是定性的表示。为定量的表示真假的程度，在模糊数学上用在闭区间[0,1]上的一个实数来表示它，叫做模糊命题的真值。例如，当真值为 1 时，命题 100%真，即“完全真”；为 0 时，命题“完全假”；为 0.2 时，则表示命题“真”的程度为 0.2。所以，模糊命题可以看做是二值逻辑中的普通命题的推广，而模糊逻辑则既是一种形式化的连续多值逻辑，又是二值逻辑的扩充，模糊命题的真值也就是模糊数学上的隶属函数值。因此，真值的运算也就是隶属函数的运算。一般情况下，模糊命题的真值为一个在 0 与 1 之间的实数 λ(即 $0\leqslant\lambda\leqslant1$)，称为它的 λ 值。模糊命题比二值逻辑中的命题更接近于人脑的思维。正是由于此，所以日本曾于 20 世纪 80 年代初，联合英、美、法、德等国，准备用十年的时间，研制一种新型的模拟人脑功能的计算机，即一种多值计算机。模糊逻辑为此研制从数学上提供了可参考的条件。

一个模糊命题 A 可用 A^* 来表示，即 A^* 就表示模糊命题 A。模糊命题真值的运算也就是模糊数学上隶属函数的运算。逻辑联结主要有：

① 逻辑并(又称析取)。设 A^* 的真值为 λ_1，B^* 的真值为 λ_2，以符号“$\vee$”表示并运算，则 $A^*\vee B^*$ 的真值定义为 $\max\{\lambda_1,\lambda_2\}$，也就是取 λ_1,λ_2 中的最大值。

② 逻辑交(又称合取)。以符号“$\wedge$”表示交运算，则 $A^*\wedge B^*$ 定义为 $\min\{\lambda_1,\lambda_2\}$，也就是取 λ_1,λ_2 中的最小值。

③ 逻辑补(否定)。A^* 的补记为 $\overline{A}^*$，$\overline{A}^*$ 的真值取 $1-\lambda_1$，$\overline{B}^*$ 的真值取 $1-\lambda_2$。

模糊命题、模糊逻辑联结和似然推理构成模糊逻辑的基础，模糊逻辑与二值逻辑不同之处，不仅在于命题不是绝对的真、假，还在于逻辑推理的似然性。似然推理是推理的一种模式，其真、假值和推理规则是模糊的而不是精确的；或者说，模糊逻辑推理具有不是很确切的特性，称这种推理为似然推理，它在医疗诊断、侦察破案、天气预报等领域都有广泛的应用。

在二值逻辑中，推理句“若…则…”（相当于英文中的“if…then…”）称为条件命题，并用 $A\rightarrow B$ 表示“若 A 则 B”。若 A^*，B^* 是两个模糊命题，则 $A^*\rightarrow B^*$ 就是似然推理。例如，设有两个模糊命题：A^*：天气晴朗；B^*：天气暖和。则“$A^*\rightarrow B^*$”就是“若天气晴朗则天气暖和”，而“$B^*\rightarrow A^*$”是“若天气暖和则天气晴朗”，这都是似然推理。

在科技竞争的时代，模糊逻辑是信息革命重要工具之一，这在美、日等国已引起高度的重视。近年来，我国模糊理论也出现了一批新的科研成果。□

3 模糊信息的研究与发展*

信息是人类认识和改造客观世界的知识源泉，人们把信息作为与材料和能源一样，并称它为第三资源。信息的概念已渗透到许多学科领域和政府部门。

信息来源于事物，来源于事物的运动，但客观事物是复杂的。有些事物的特征明显，可以有明确的界限，这类事物叫清晰事物；但更多的事物的界限是不分明的，模棱两可，含混不清，在人脑里反映出不确定的数量概念，叫做模糊事物。例如"热"与"不热"，究竟多少摄氏度才算"热"或"不热"，并没有一个公认的定量标准或界限。因为气温的变化是逐渐的，连续不断的，不存在一个突然的转折，致使"热"与"不热"的分界线模糊不清。"热"与"不热"作为信息，叫做模糊信息。在疾病的诊断中，医生尝试应用模糊信息方法，使诊断的准确性能有所提高。任何疾病都要产生信息，而模糊信息在医学中，尤其是在诊断医学中是普遍存在的。如心率慢、胆固醇高等都是不明确的模糊的疾病信息，因此就给医学工作者提出了应用信息与模糊方法来探索和诊断疾病的问题。有人提出，针灸和麻醉都是模糊信息疗法。例如针灸是用各种刺激手段向人体输入信息，伴随一点点能量，作为信息的载体。针灸所刺激的穴位是选取最佳的信息输入点，即经络反应最强烈的穴位，向失调的经络输入反馈信息，恢复和促进人的肌体的经络调节作用，使肌体恢复健康。这个过程使用了许多模糊方法，如伴随"一点点能量"、"最佳穴位"、"反应最强烈"

---

*本文原载于《自动化博览》1991 年第 1 期。

以及“失调”等。显然，在模糊控制中使用的信息，就是模糊信息。

模糊性产生于由量变到质变的连续过渡。由于划分的不确定性，造成了元素对集合隶属关系的不确定性。如一年四季之间的时间划分，现在、过去、未来之间的划分等，都无严格的界限，它们之间不存在不可逾越的鸿沟。由于这类事物的界限是模糊的，所以关于这类事物的信息就是模糊信息。

所谓模糊信息是指现实世界中的一类事物的信息，由属于此类到不属于此类的界限是渐近的、模糊的、不确定的，即指消息本身含义模糊不确定的信息。

进入 20 世纪 60 年代，随着人们社会实践的深入和科学的发展，所处理和研究的领域已进入复杂系统、大系统和超大系统，如生态系统、宇航系统、人脑系统等。这类系统不仅结构与功能复杂，涉及大量的参数与变量，而且内部关系非常错综复杂，具有模糊性，用传统的数学方法难以精确地描述，也无法真实地反映人脑里的思维活动。以传统数学及二值逻辑为基础而建立的电子计算机不具备像人脑那样灵活处理模糊事物的能力，这对计算机智能的发展是一个极大的障碍。为了能把人的自然语言作为算法语言而编入程序，让计算机能够描述和处理事物的模糊性，就必须建立起相应的能够描述和处理模糊量、模糊信息及其关系的数学方法。1965 年，美国控制论学家扎德总结了模糊现象的规律，发表了题为《模糊集合》的论文，创立了模糊数学这一新的研究领域。今天，模糊已不再是人们难以接受的概念，模糊数学对许多人来说，也已不再是很陌生的科学了。

1972 年，意大利的德路卡等人发表了《在模糊集合论的基础上的非概率熵的定义》一文，研究了纯模糊性所引入的不确定性，将仙农的概率熵的概念推广到模糊集合上，建立了模糊集合上的非概率的模糊熵的定义。这是与仙农的信息量不同的另一种信息度量，用非概率、非统计的方法来建立信息的一种在性质上全新的模型，着重从质的方面对信息进行定性研究。这对仙农的信息论的定量分析来说，是一个重大突破，从而导致了模糊信息与模糊熵的研究。顺便指出，概率论或统计学中的不确定性与模糊数学中的不确定性有着本质上的不同。概率论中的不确定性源于随机性，是指事件的发生与否而言，而事件本身的含义是确定的，只是由于条件的不充分，致使事件的发生与否有多种可能，即所谓不确定性。模糊数学中的不确定性源于模糊性。所谓模糊性是指元素对集合的隶属关系而言，事件本身的含义是不确定的，但事件的发生与否是可以确定的，因而元素（事件）对集合的隶属关系是不确定的。但在某些事物内部，有时随机性与模糊性是共存的。例如天气预报“本月下旬的最高温度大约是 35℃”，这里的“大约”是对温度的一种不确定性的描述，即概念的模糊性；同时它又是对出现 35℃的一种不确定性的描述，即事件的随机性。

熵最早产生于物理学领域，原是热力学中的一个概念，表征物质系统的热学状态，以后又渗透到其他领域，得到广泛的应用。例如在统计力学中，熵是用来描述分子运动的无序性（即无规则性）的一种度量；在模糊数学中，利用熵来描写模糊程度，是模糊集合所含模糊性大小的一种量度；在概率论的信息论中，熵是表述事件的不确定性的一种度量，即剩余信息量大小的一种度量。

在我国，也有人对模糊信息进行了研究，并把仙农信息推广到模糊情形，提出了模糊熵、广义信息论、广义信息源和广义熵等问题。广义信息源一般是随机的，又是模糊的，即具有双重不定性。比如研究某一模糊事件是否发生，既要考虑事件本身含义模糊性上的不定性，又要考虑事件是否发生随机性上的不定性。无论是从单一的随机性还是单一的模糊性，都不能对其进行全面刻画，只有同时考虑随机性和模糊性这两个方面的不确定性才是比较恰当的。这种具有随机性和模糊性的双重不定性的信息源称为广义信息源，其熵称为广义熵，其既包含表达信息源随机性的仙农概率熵，又包含表达信息源模糊性的模糊熵。因此研究和处理广义熵既需要概率论，又需要模糊数学这两种不同的数学工具。

模糊数学是信息革命的重要工具，已引起美、日等国各界的重视。例如，日本 30 家大公司集资 30 亿日元筹办"国际模糊工程研究所"。我国虽然是仅次于美、日两国的模糊数学大国，但是重大的突破性理论和应用还没有见到。严格说，理论的体系还远未形成，在模糊集合论或模糊数学基础上建立起来的模糊信息和模糊熵以及模糊决策等问题，目前还处在发展的初期，许多定义和概念并不统一，至于系统的模糊信息论仅仅是处于萌发阶段。为了发展和完善我国模糊数学，并参与国际竞争，一些模糊数学工作者提出"把我国模糊数学的研究重点集中到模糊信息处理与机器智能的前沿阵地上来"。北京师范大学模糊数学专家汪培庄教授组织 29 个单位的 155 位模糊数学工作者，向国家自然科学基金委员会申请重大科研项目"模糊信息处理与机器智能"。基金委员会于 1988 年组织专家论证，正式批准并下达任务书，拨发研究经费 135 万元。该项目共分 8 个子课题，较全面地涉及模糊数学理论在模糊信息处理和机器智能方面的应用，既包括基础理论与应用，也包括计算机软件和硬件。这是我国模糊数学截至目前规模最大的科研项目。□

4 模糊推理与模糊控制器及其发展*

一 模糊语言与模糊推理

人类生存环境，基本上是一个模糊环境。人们在日常活动中，经常接触模糊事物，接受模糊信息，随时对模糊事物进行识别，作出决策。所谓模糊性，是相对于清晰性而言，主要是指客观事物差异的“不分明性”，也就是“亦此亦彼性”。没有明确界线的概念和现象，叫做模糊概念和模糊现象。描述这种概念和现象的语言，叫做模糊语言。人们在日常生活中使用的语言称为“自然语言”，它用字或词作为符号，按一定的规则（语法）排列，就可以构成句子，表达人们的思想。这些句子具有模糊性或多义性。由于人脑思维的特性，对于这些句子不仅理解很快，还会产生美感和幽默感，使人们感到舒适和兴奋。但是，要使计算机理解和接受人类的自然语言却很困难，因为现在的计算机是建立在二值逻辑的基础上，二值逻辑对任何命题的判断“非真即假”或“非此即彼”。但人脑的思维活动，并不是绝对的抉择，而是在“真”、“假”和“是”、“非”之间存在着过渡状态。处于过渡中的思维活动具有模糊性，反映出来就是所谓模糊信息，按精确数学的原理是无法描述的，但在日常生活中，人们却能根据这些模糊信息作出判断。通常使用一些符号和法则编制的“计算机语言”，不允许有模糊性

* 本文原载于《自然辩证法研究》1993 年第 7 期。

和多义性，只是一种形式上的语言，称为“形式语言”。只按照“形式语言”工作的计算机不具有灵活性，不能模拟人脑的思维，不适合在具有“模糊性”的控制系统中使用。因此，要使自然语言能直接进入计算机程序，就必须在形式语言中渗入自然语言。这种渗入了自然语言的形式语言，就是模糊语言。它是语言学和数学结合的产物，是模糊自动控制所采用的基本语言。

在模糊自动控制中，控制规则一般是以模糊条件语句形式出现的。常见的模糊条件语句是形如“如果…那么…”“若…则…否则…”的语句。如：“如果反应炉温度很高，且迅速上升，那么将冷却水稍微增加一些。”“如果 x 略小，则 y 略大，否则，y 不很大。”

设系统的输入是论域 U 上的模糊子集 $\underset{\sim}{A}$，输出是论域 V 上的模糊子集 $\underset{\sim}{B}$，则 $\underset{\sim}{A}$，$\underset{\sim}{B}$ 间的函数关系为“若 $\underset{\sim}{A}$ 则 $\underset{\sim}{B}$”形式的条件语句，它可以表示为模糊关系 $\underset{\sim}{R}=\underset{\sim}{A}^T\times\underset{\sim}{B}$，其中 $\underset{\sim}{A}^T$ 为输入列向量。在模糊自动控制中经常遇到的是多段条件语句。例如，“若 $\underset{\sim}{A}_1$ 则 $\underset{\sim}{B}_1$，…，若 $\underset{\sim}{A}_n$ 则 $\underset{\sim}{B}_n$。其中 $\underset{\sim}{A}_1$，…，$\underset{\sim}{A}_n$ 为论域 U 上的模糊子集，$\underset{\sim}{B}_1$，…，$\underset{\sim}{B}_n$ 为论域 V 上的模糊子集，则多段条件语句可以用 U 到 V 上的一个模糊关系 $\underset{\sim}{R}$ 来表示。”

模糊推理又叫近似推理，顾名思义，这是模拟人的日常推理的一种近似推理。

经典的二值逻辑是人们传统推理的基础，这种推理对任何命题的判断“非真即假”或“非此即彼”。这种精确推理的常用形式是所谓“三段论法”（即经过三个阶段得出结论的推理方法），但是在很多情况下很难绝对判断一个命题的“真”或“假”。如对含有模糊概念的陈述句，即所谓模糊命题，它就不取绝对的“真”、“假”，不符合二值逻辑，而是取 0 与 1 之间的一个实数来表示它的真值。因此，它是需要用多值逻辑来描述的。模糊命题的真、假值也就是模糊数学上的隶属函数值，真、假的运算也就是隶属函数的运算。例如，设模糊命题 $\underset{\sim}{A}$，$\underset{\sim}{B}$ 的真值分别为 λ_1，λ_2，λ_1，λ_2 均在 0 和 1 之间取值，则 $\underset{\sim}{A}\vee\underset{\sim}{B}$ 的真值为 $\lambda_1\vee\lambda_2\triangleq\max\{\lambda_1,\lambda_2\}$，$\underset{\sim}{A}\wedge\underset{\sim}{B}$ 的真值为 $\lambda_1\wedge\lambda_2\triangleq\min\{\lambda_1,\lambda_2\}$，符号“$\vee$”表示取最大值运算，“$\wedge$”表示取最小值运算，“$\triangleq$”表示定义。模糊命题比二值逻辑更接过于人脑的思维，它是模糊逻辑的基础。模糊逻辑与二值逻辑的不同之处，不仅在于命题不是取绝对的真、假值，还在于逻辑推理的“模糊性”，即推理具有不很确切的特性。其实，人们日常思考问题时，也常常有这样的推理方法。如以“若 x 小，则 y 大”为依据，“如果 x 很小”，就判断“y 很大”；“如果 x 略小”，就判断“y 略大”。这种推理过程，可以看做一种模糊变换，它将“x 很小”与“x 略小”的真

域 $\underset{\sim}{A}_1$ 与 $\underset{\sim}{A}_2$ 分别变换到“y 很大”与“y 略大”的真域 $\underset{\sim}{B}_1$ 与 $\underset{\sim}{B}_2$，这种推理方法称为模糊推理。它是以模糊判断为前提，使用模糊推理规则，推出以模糊判断为结论的推理，其中关键之处是推理规则具有模糊性。模糊推理是二值逻辑推理在另一种意义下的推广。运用扎德的模糊集合理论，可以对模糊推理问题作定量的分析。

二　模糊控制

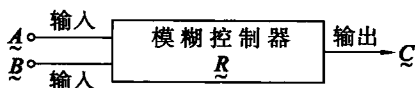
近年来，用计算机来控制生产过程在许多领域已得到了成功的应用。这种自动控制(包括经典控制和现代控制)的目的就是借助于计算机使生产经常处于最优状态。它具有形式化、数量化的特点。因此，研究和设计自动控制的第一步，就是要建立计算机能够解算而又与生产相适应的精确数学模型(或为传递函数，或为状态方程)。可是现代控制论的研究对象极为复杂，对于那些时变的、非线性的和生产过程复杂的系统，最困难的问题之一就是数学模型不易求得，而且数学模型也不是一劳永逸的，会随生产条件的改变而改变。此外，在设计控制系统时，还必须考虑工程上的具体问题。例如，简单性、可靠性、成本低等等。因此用计算机控制的生产过程遇到了一系列的困难，不能用计算机控制的生产过程，由熟练的专业工人，凭他的丰富经验却控制得非常好。例如，有这样一条指令：“如果反应炉中温度很高，且迅速上升，那么将冷却水稍增加些。”这里，“很高”、“迅速上升”和“稍增加些”都是模糊信息，工人能灵活应用，控制得很好，而由计算机指挥的炼钢炉，就不能执行由模糊信息构成的指令。这就给人们一个启发，即绕过建立精确数学模型这一关，把人脑的直觉和技师的经验加以整理与处理，归纳成一组条件语句，然后用模糊数学的工具加以数量化，并用模糊逻辑、模糊语言给出模糊算法，使控制器能根据人的经验，模仿人的控制策略、控制指令，实现对系统的最优控制(可称为模糊最优控制)。上面所说的那种控制器(即算式)就是模糊控制器。这种控制器与维纳所建立的信息反馈式的控制器在基本原理上是有所不同的。这种新型的控制器是模糊数学与控制论相结合的产物，是传统控制器的发展。

关于模糊控制器的设计可以简要归纳为以下三个方面：

① 把各语言变量化为某个论域上的模糊子集。一般是通过计算时的误差及误差变化(或误差变化率)，并将误差变化分类定级，给出在各种组合情况下的隶属度，使各种变量模糊化。

② 设计模糊算法，计算出模糊控制量。模糊控制算法通常是应用模糊条件语句或模糊推理来表达。例如：若 $\underset{\sim}{P}$ 则 $\underset{\sim}{Q}$(if…then…)；若 $\underset{\sim}{P}$ 或 $\underset{\sim}{Q}$ 则 $\underset{\sim}{S}$(if…or…then…)；若 $\underset{\sim}{A}$ 则 $\underset{\sim}{B}$ 否则 $\underset{\sim}{C}$(if…then…else…)；若 $\underset{\sim}{A}$ 且 $\underset{\sim}{B}$ 则 $\underset{\sim}{C}$(if…and…then…)等等。每一条语句都对应一个模糊关系 $\underset{\sim}{R_i}$，各模糊关系之并($\bigcup_i \underset{\sim}{R_i}$)即总的模糊关系。通过每一条规则可算出对应的控制量，各控制量之并就是总的控制量，它是一个模糊子集。

③ 模糊判决方法的设计。一般的模糊控制方式如下图所示：



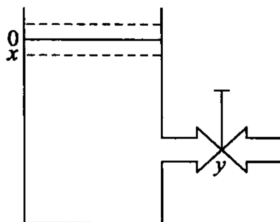
模糊控制器可以看成是一个模糊关系 $\underset{\sim}{R}$，输入通常为二维的模糊集 $\underset{\sim}{A}$ 和 $\underset{\sim}{B}$，输出是一维的模糊集 $\underset{\sim}{C}$，输入输出的变换通过模糊关系 $\underset{\sim}{R}$ 来完成。模糊控制器的输出是一个模糊子集，它包含了控制量的各种信息。若受控对象只能接受一个控制量，那么，就需要从输出的模糊子集中判决出这一个控制量。所谓判决，就是设计一个由模糊集合到普通集合的映射。模糊判决的方法很多，主要有三种：最大隶属原则，中位数判决，加权平均判决。每一种方法都有各自的优缺点。例如，最大隶属原则，在模糊控制器输出的模糊集中，取隶属度大的那个论域元素 μ^* 作为控制器的判决输出，它只考虑了主要信息，抛弃了其他次要的信息，因此缺乏全面的考虑。为此可采用中位数判决，即将隶属函数曲线与横坐标所围成的面积平均分成两部分，所对应论域元素 μ^* 作为输出判决。这种判决方法能概括更多的信息，但没有突出主要信息。加权平均判决考虑到了各个因素不同的重要程度，按重要程度不同乘以相应的权系数，然后相加，取其和作为输出。但如何选择合适的加权系数却是个难题，而这种方法的关键正是加权系数的选择。所以在实际问题中，要根据具体情况选择一种适当的方法。

三　模糊控制的应用

1974 至 1976 年，英国工程师马丹尼首先把近似推理和模糊条件语句组成模糊控制器，并应用于蒸汽机和锅炉的控制。后来又有人把模糊控制器用于搅拌池的温度控制、压力容器控制、水泥窑的工艺控制、热交换过程控制。1977 年，佩比斯(Pappis)和马丹尼又将模糊控制方法用于交叉路口的交通控制并获得了较好的效果。1982 年，日本山形大学中津千男等人进一步研究了城市十字

路口交通的模糊控制问题，实现了计算机控制。它的算法语言比较简单，采用微型计算机即可实现。利用模糊理论研制的模糊逻辑计算机一问世，便受到人们的青睐，把装置安装在地下铁道系统，列车的加速和刹车就变得非常平稳。日本松下电气公司研制成功的"模糊洗衣机"，具有 600 多种周期组合，唯一的控制是一个启动钮，其他的事如判断载荷的多少和污秽程度，确定最佳的洗衣时间和用水量，以取得最佳的洗涤效果等，全都交给传感器和模糊控制器去做。宾馆、旅店客人来来往往，由于每人看电视的主观标准不一样，对颜色、亮度和对比度总要调来调去，而用模糊技术安装的电视机就不用客人忙乎一阵了。这种电视机具有主动信号校正的功能，对颜色、亮度和对比度每隔$\frac{1}{60}$秒比对一次，它将画面上的 248 个点同存储在机内的 40 幅"完美"的画面的图像库进行比较，寻找相似而不必是相同的条件。

以下介绍模糊理论中常提到的一个典型例子——水位的模糊控制，以说明模糊控制的要点。下图为一水箱，具有可变水位 x，y 是调整阀门，"负向开"代表往外排水，"正向开"代表往内注水，0 为标准水位。要设计一个模糊控制器，并通过调节阀将水位稳定在标准点"0"附近。



假定对箱内水位变化原因不能精确掌握，人们往往是凭经验去控制水位，这些经验可以归纳为一些粗略的用语言描述的控制规则：

① 若 x 很大于 0(叫正大，记作 PB)，则 y 大量排水(叫负大，记作 NB)；② 若x 稍大于 0(叫正小，记作 PS)，则 y 稍许排水(叫负小，记作 NS)；③ 若 $x=0$，则 y 保持不动(记作 $y=0$)；④ 若 x 稍小于 0(叫负小，记作 NS)，则 y 稍许注水(叫正小，记作 PS)；⑤ 若 x 很小于 0(叫负大，记作 NB)，则 y 大量注水(叫正大，记作 PB)。

观察量是实际水位与标准水位偏差 e，控制量是阀门转盘角度的变化 μ(逆时针转为正(注水)，顺时针为负(排水))。在实际工作中，e 和 μ 都是连续变化的，把 e 和 μ 的数值离散化，分为 5 档，分别为正大、正小、0、负小、负大。于是，控制指令为以下一组模糊条件语句：

① 若 e 正大，则 μ 负大；② 若 e 正小，则 μ 负小；③ 若 e 为 0，则 μ 为 0；

④ 若e负小，则μ正小；⑤ 若e负大，则μ正大。

用适当的模糊集合表示正大、正小、0、负小、负大，用适当的模糊关系表示各个模糊指令，于是便形成了控制系统的一种模糊的定量模型。设计适当的控制器，即可实现对水位的自动控制。

四　模糊控制器的发展——自组织模糊控制器

近年来，在模糊控制的研究中，又出现了一种自组织模糊控制器，这种控制器是一般的模糊控制器的发展。

我们知道，模糊自动控制的规则基于人们对控制经验和信息的总结与归纳，但是客观事物是复杂的，对于有些生产过程，特别是存在随机干扰的情况下，人们对控制经验和信息往往缺乏认识或认识不清，总结不出经验，因此对模糊控制的规则就很难制订出来。即使对一些控制规则比较完善的系统，由于环境等多种因素的影响，致使原有的规则不能运用于变化了的实际过程，这就迫使人们不得不去研究具有能在运行中自动修改、完善和调整模糊控制的功能，以使控制系统的性能不断改善，直至达到预定的效果。这种具有自组织功能的模糊控制器就称为自组织模糊控制器。它在原模糊控制器的基础上附加了性能测量、控制量校正及控制规则修正三个部分，增加这些部分是为了在运行过程中，自动调整和修改控制规则，从而达到改变控制系统性能的目的。

自组织模糊控制器是在一般的自组织控制原理和模糊控制器的基础上，1979 年由英国学者普鲁塞克和马丹尼提出来的一项研究成果。关于自组织模糊控制器的研究还很不成熟，但这是一个很诱人的研究领域。□

5 模糊模式识别概说*

一 什么是模糊模式识别

模式识别也称图形识别，简单地说，就是把一种研究对象，根据它的某些特征进行识别并分类的一门综合性新兴技术。它所研究的对象，如文字、图像、产品或物体等，概括起来统称为模式。比如，邮局每天识别大量的信函上的地区或编码，并按此分类，以便送到各个相应的地区；人们从金属的敲击声中辨认该金属的性质与成分，确认其真假；从遥感图片中辨别农作物以及判定它的长势，并预测它的收获量等，都是模式识别。随着电子技术及计算机技术的突飞猛进，模式识别已进入一个利用电子计算机进行识别的崭新时代，但现实世界中的文字、图像以及其他物体，很多是不清晰的，如“潦草的手写字”、“大个子与矮个子”、“大胡子与小胡子”等，这些没有确定的标准或界限的现象称为模糊现象，当识别的对象为模糊现象时就称为模糊模式识别。如一封有若干字迹被雨水冲洗掉的信，人能从上下文探寻到其真实含义；公安人员或法医为了破案，对嫌疑犯的指纹进行鉴别；中医对病人诊病等，都是模糊模式识别。上述探寻信的真实含义、对嫌疑犯的指纹进行鉴别的过程等，都是模糊模式识别的过程。目前的电子计算机，就计算速度而言，一秒钟的运算可达千万次以上，但

*本文原载于《自动化博览》1989年第3期。

对模糊模式的识别和分类，却爱莫能助。如看电视时，要把图像调得更清楚些，这对一个五六岁的儿童都是轻而易举的事，但计算机无能为力。如何使计算机吸取人脑识别和判决模糊模式的特点，使其能仿效人脑进行模糊识别与判决，就成为科技界竞相研究的热门课题。但现在的电子计算机的理论是建立在二值逻辑的基础之上的，对任何命题，非 0 即 1（即非假即真），仅仅这两个值是绝对不能真实地反映千变万化活动的人脑思维的。美国自动控制论学家、加利福尼亚大学教授扎德认识到电子计算机的上述缺陷，于 1965 年发表了题为《模糊集合》的开创性论文，把经典集合中的特征函数仅取 0 与 1 这两个值的二值逻辑关系推广到可取 0 与 1 之间的任何实数值的多值逻辑关系，并把此函数叫做论域 X 上的集合 A 的隶属函数，记为 $\mu_A(x)$，A 叫做论域 X 上的模糊子集，$\mu_A(x)$称为 x 对 A 的隶属度。

二　模糊模式识别的方法

1. 最大隶属原则

设论域 X 上有 n 个模糊子集 $A_1,A_2,\cdots,A_n$，它们可看成是模糊模式。若对于元素 $x_0\in X$ 满足

$$\mu_{A_i}(x_0)=\max\{\mu_{A_1}(x_0),\mu_{A_2}(x_0),\cdots,\mu_{A_n}(x_0)\},$$

则判定 x_0 属于 A_i，因为 x_0 隶属于 A_i 的程度最高。

这种直接由计算元素的隶属度来判断具体对象归属的方法又称为模糊模式识别的直接方法，可用来识别明确的对象 $x_0\in X$ 的归属类别。用此法识别模式归属的关键是如何合理地确定隶属函数，但其确定的方法至今仍无规律可循，主要是靠经验和技巧。

2. 择近原则

在现在生活中，不仅要考虑一个元素对模糊集合的隶属关系，还要考虑两个模糊集合彼此靠近的程度。两个模糊集合彼此靠近的程度叫贴近度。例如在工业生产中，未经处理的铸件毛坯零件是模糊的，摄像获得的图像更加模糊，要从图像中识别所需零件，就得用择近原则。再如文字识别，输入的字与字的模型都是平面上格子点的模糊子集，也属于这一种模式识别。

定义　若 A 与 B 都为论域 X 上的模糊子集，A 与 B 的贴近度记为(A,B)，定义为

$$(A,B)=\frac{A\circ B+(1-A\odot B)}{2}。$$

贴近度是[0,1]上的一个数，它在模糊模式识别的问题上至为重要。

择近原则 设 $A_1, A_2, \cdots, A_n$ 为论域 X 上的 n 个模糊子集，被识别的对象 B 是论域 X 上的另一个模糊子集。对于 B，若有 $i(i \in 1, 2, \cdots, n)$ 使

$$(B, A_i) = \max_{1 \leqslant j \leqslant n} \{(B, A_j)\},$$

则认为 B 与 A_i 最贴近，把 B 归属于 A_i。

从系统辨识的角度来理解，如果以模糊子集 $A_i, A_2, \cdots, A_n$ 表示 n 个模型，而模糊子集 B 与 A_i 最贴近，就把 B 说成是相对符合于模型 A_i 的。□

6 模糊综合评判

一　综合评判

在生产、科研或日常生活中，经常需要对受某种因素影响的事物进行评估，再通过一定的数学方法把评估结果定量化，以决定优劣，这就是所谓的综合评判。如对某一项科技成果，以技术水平、经济效益、需要程度、投资需要量等因素进行评估；又如对某服装店制作的衣服，就其花色、式样、价格、耐久度和舒适度等进行评估；诸如此类，不胜枚举。设影响事物的因素有 n 个，且各因素对该事物的影响是一样的，第 i 个因素的评分为 S_i，总的评分为 W，则

$$W=\sum_{i=1}^{n} S_i。$$

在实际问题中，往往侧重某些因素。例如某大学招收工科硕士研究生，考试课程为英语、高等数学、力学和机械设计，学校就这些课程进行综合权衡之后认为：高等数学和力学较之另外两门课程格外重要，因此规定在计算总平均分数时各占 30%，而英语和机械设计各占 20%。这些百分比被称之为权（重）。某考生四门课的考试成绩分别是：英语 70 分，高等数学 80 分，力学 60 分，机械设计 85 分，则该考生的实际综合得分为 $W=70\times0.2+80\times0.3+60\times0.3+85\times0.2=73$。

每一门课的得分，实质上是对单因素的一个评判。加权平均所得的分数，实质上就是综合评判的结果。一般

地，设 $a_i(i=1,2,\cdots,n)$ 是第 i 个因素所占的权重（a_i 是根据实际情况主观决定的，且规定 $\sum_{i=1}^{n} a_i=1$），则加权平均分可表示为

$$W=\sum_{i=1}^{n} a_i S_i,$$

式中 $S_i(i=1,2,\cdots,n)$ 是对第 i 个因素的评分。

加权平均分和总评分起着相同的作用，但它能比计算总分更加合理地作出评判，因为计算总分时把每一个因素对被评事物的影响看成是等同的，然而，这对实际问题来说，却未必是合理的。

二　模糊综合评判

综合评判是在确定的情况下作出的判决或决策，而且对每一个因素都有一个确定的评分，但实际上对影响事物的很多因素却难以用简单的分数来评判。例如对某品牌的彩色电视机，如何来评价其图像是清晰或不清晰，就不能用简单的分数来回答，因为清晰或不清晰没有一个公认的标准或界限来加以判断。在经典的集合论里，论域 Z 中的一个元素 x 与已知集合 A 的关系是 $x\in A$ 或 $x\notin A$，二者必居也只居其一，绝对不能模棱两可。用特征函数 $\mu_A(x)$ 表示，就是

$$\mu_A(x)=\begin{cases}1, x\in A;\\0, x\notin A。\end{cases}$$

经典的集合论处理模糊现象遇到了不可克服的困难。1965 年，美国自动控制学家、加利福尼亚大学教授扎德发表了题为《模糊集合》的论文，把经典集合论中特征函数 $\mu_A(x)$ 只取 0 与 1 这两个值推广到 $\mu_A(x)$ 可取 0 与 1 之间的任何实数值，把在闭区间[0,1]上取值的这个函数叫做集合 $\underset{\sim}{A}$ 的隶属函数，记为 $\mu_{\underset{\sim}{A}}$，从而 $0\leqslant\mu_{\underset{\sim}{A}}(x)\leqslant 1$。$\underset{\sim}{A}$ 叫做论域上的模糊子集。$\mu_{\underset{\sim}{A}}(x)$ 称为 x 关于 $\underset{\sim}{A}$ 的隶属度。$\mu_{\underset{\sim}{A}}(x)$ 的值越接近于 1，就表示 x 隶属于 $\underset{\sim}{A}$ 的程度越高；$\mu_{\underset{\sim}{A}}(x)$ 的值越接近于 0，就表示 x 隶属于 $\underset{\sim}{A}$ 的程度越低。当 $\mu_{\underset{\sim}{A}}(x)=1$ 时，x 完全（百分之百）属于 $\underset{\sim}{A}$；当 $\mu_{\underset{\sim}{A}}(x)=0$ 时，x 完全不属于 $\underset{\sim}{A}$。因此，隶属度表示了 x 隶属于 $\underset{\sim}{A}$ 的程度。模糊集合的创立，标志着模糊数学的诞生。隶属函数是一个构造函数，它是根据实际情况和经验主观构造的。不言而喻，隶属函数是描述模糊性现象的一个关键。

本文所用到的模糊数学的主要运算是：

并:两个模糊集 $\underset{\sim}{A}$ 和 $\underset{\sim}{B}$ 的并也是一个模糊集,记为 $\underset{\sim}{A}\cup\underset{\sim}{B}$,定义为包括 $\underset{\sim}{A}$ 和 $\underset{\sim}{B}$ 的最小模糊集。并集 $\underset{\sim}{A}\cup\underset{\sim}{B}$ 的隶属函数记为 $\mu_{\underset{\sim}{A}\cup\underset{\sim}{B}}(x)$,由

$$\mu_{\underset{\sim}{A}\cup\underset{\sim}{B}}(x)=\max\{\mu_{\underset{\sim}{A}}(x),\mu_{\underset{\sim}{B}}(x)\}\triangleq\mu_{\underset{\sim}{A}}(x)\vee\mu_{\underset{\sim}{B}}(x),\forall x\in\underset{\sim}{A}\cup\underset{\sim}{B}$$

所确定。

交:模糊集 $\underset{\sim}{A}$ 和 $\underset{\sim}{B}$ 的交集记为 $\underset{\sim}{A}\cap\underset{\sim}{B}$,定义为同是这两个集合的子集的最大模糊集。集合 $\underset{\sim}{A}\cap\underset{\sim}{B}$ 的隶属函数记为 $\mu_{\underset{\sim}{A}\cap\underset{\sim}{B}}(x)$,由

$$\mu_{\underset{\sim}{A}\cap\underset{\sim}{B}}(x)=\min\{\mu_{\underset{\sim}{A}}(x),\mu_{\underset{\sim}{B}}(x)\}\triangleq\mu_{\underset{\sim}{A}}(x)\wedge\mu_{\underset{\sim}{B}}(x),\forall x\in\underset{\sim}{A}\cap\underset{\sim}{B}$$

所确定。

模糊矩阵:元素仅在 0 与 1 之间取值的矩阵叫模糊矩阵。其一般形式为

$$\underset{\sim}{\boldsymbol{A}}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

也可简记为 $\underset{\sim}{\boldsymbol{A}}=(a_{ij})_{m\times n}$,其中 $0\leqslant a_{ij}\leqslant 1,1\leqslant i\leqslant m,1\leqslant j\leqslant n$。

模糊矩阵代表了一种模糊关系,故又称为模糊关系矩阵。

设模糊矩阵为 $\underset{\sim}{\boldsymbol{A}}=(a_{ij})_{m\times n}$,称 $(1-a_{ij})_{m\times n}$ 为 $\underset{\sim}{\boldsymbol{A}}$ 的补矩阵,记为 $\overline{\underset{\sim}{\boldsymbol{A}}}$。

模糊矩阵 $\underset{\sim}{\boldsymbol{A}}$ 和 $\underset{\sim}{\boldsymbol{B}}$ 的乘法记为 $\underset{\sim}{\boldsymbol{A}}\circ\underset{\sim}{\boldsymbol{B}}$,又称为 $\underset{\sim}{\boldsymbol{A}}$ 与 $\underset{\sim}{\boldsymbol{B}}$ 的合成。它与普通矩阵的乘法类似,不同的是并非先两项相乘后相加,而是先取小后取大。记 $\underset{\sim}{\boldsymbol{A}}\circ\underset{\sim}{\boldsymbol{B}}=\underset{\sim}{\boldsymbol{C}}$,则 $\underset{\sim}{\boldsymbol{C}}$ 的元素为

$$c_{ij}=\max_{k}\{\min\{a_{ik},b_{kj}\}\}=\bigvee_{k}(a_{ik}\wedge b_{kj})。$$

由于被评价的对象及影响对象的因素很多是模糊的,因而有所谓模糊综合评判。例如,商店里出售某种服装是否为顾客所喜欢,往往牵涉到好几个因素,如花色、样式、耐久度和价格等,而顾客是否喜欢这种服装和每一个因素都有关系,然而喜欢和因素又都是模糊的。如何来评判一件服装的优劣,这就是一个模糊综合评判问题。模糊综合评判的数学模型可归纳如下:

记已知影响事物的因素集合为 $X=\{x_1,x_2,\cdots,x_n\}$,对事物的评语集合为 $V=\{\underset{\sim}{v_1},\underset{\sim}{v_2},\cdots,\underset{\sim}{v_n}\}$。

又设应用统计试验或专家评分等方法建立各因素间的权重分配为 X 上的模糊子集,并写成矩阵形式

$$\underset{\sim}{\boldsymbol{A}}=(a_1,a_2,\cdots,a_n),$$

式中 a_i 为第 i 个因素 x_i 所对应的权重,规定

$$\sum_{i=1}^{n} a_i = 1。$$

对第 i 个因素的单因素模糊评判为 V 上的模糊子集写为矩阵形式

$$\underset{\sim}{\boldsymbol{R}}_i = (r_{i1}, r_{i2}, \cdots, r_{in})。$$

把从 1 到 n 个单因素模糊评判结果作为行，构成一个模糊关系矩阵（有时也叫做单因素模糊评判矩阵）

$$\underset{\sim}{\boldsymbol{R}} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix},$$

则对该评判对象的模糊综合评判 $\underset{\sim}{B}$ 是 V 上的模糊子集（矩阵形式）

$$\underset{\sim}{\boldsymbol{B}} = \underset{\sim}{\boldsymbol{A}} \circ \underset{\sim}{\boldsymbol{R}}。$$

再借助于合成运算，就可得出模糊综合评判的结果。

三　模糊综合评判在评价教师课堂讲授效果上的应用

如何评价教师课堂讲授效果，这是当前教学改革和教学管理上的一个重要课题。过去对教师的课堂讲授效果是通过对学生的平时反映和对这门课程的考试成绩来考核的，这确实可以了解到一些情况，但作为评价教师课堂讲授效果好坏的依据，就似乎不够充分。教师课堂教学的好坏是一个模糊概念，它不仅涉及到所教课程的知识，还涉及到教育学、心理学、语言学等学科，而讲课的优劣对教育质量的提高有着决定性的影响。因此，正确评价教师课堂讲授的效果，对教学进行数字化、定量化的综合评判，不但有助于教师总结经验，提高课堂的讲授艺术，同时对教师的考核、定级也提供了数量依据。

课堂教学是一个多因素的复杂智力劳动，涉及的因素很多。如讲授知识的深广度、学生对知识的接受情况、学生能力的培养、教学方法的启发性以及教师的语言等等，而这些因素也都是模糊的。因此，如何评价教师课堂教学效果的好坏，这是一个模糊综合评判问题，用模糊数学的方法来处理或评价才较为科学。因此，可以邀请若干位教学经验丰富的专家，通过连续若干次听课和搜集学生意见，进行综合评判。

设给出评价集为

$$V = \{优秀，良好，一般，差\},$$

或记为 $$V=\{x_1,x_2,x_3,x_4\},$$
其中 x_1,x_2,x_3,x_4 分别代表优秀、良好、一般、差。

因素集为

$X=\{$知识的深广度，学生的意见，对学生能力的培养，启发性，教师语言$\}$。

首先，对各个单独因素进行模糊评判。对“知识的深广度”这一因素，20%的专家认为是优秀，50%的专家认为是良好，30%的专家认为一般，没有专家认为是差的。这一评判结果可用一个模糊集

$$\underset{\sim}{R}_1=\frac{0.2}{x_1}+\frac{0.5}{x_2}+\frac{0.3}{x_3}+\frac{0}{x_4}$$

来表示（上式是扎德对模糊集的一种表示方法，不能认为是通常的分数），也可表示成一个行矩阵

$$\underset{\sim}{R}_1=(0.2,0.5,0.3,0)。$$

类似地，假定专家对其他四个因素即“学生的意见”、“对学生能力的培养”、“启发性”、“教师语言”进行评价，其评价结果依次为

$$\underset{\sim}{R}_2=(0.1,0.3,0.5,0.1),$$
$$\underset{\sim}{R}_3=(0,0.4,0.5,0.1),$$
$$\underset{\sim}{R}_4=(0,0.1,0.6,0.3),$$
$$\underset{\sim}{R}_5=(0.5,0.3,0.2,0)。$$

以它们为行构成一个模糊关系矩阵

$$\underset{\sim}{R}=\begin{pmatrix}0.2 & 0.5 & 0.3 & 0\\ 0.1 & 0.3 & 0.5 & 0.1\\ 0 & 0.4 & 0.5 & 0.1\\ 0 & 0.1 & 0.6 & 0.3\\ 0.5 & 0.3 & 0.2 & 0\end{pmatrix}。$$

这就是对教学效果的单因素模糊评判矩阵。其中各个因素对教学效果的影响程度显然是不一样的。设应用统计试验或专家评分法对上述五个因素的权重分配依次是 0.35,0.2,0.2,0.15,0.1，写成矩阵形式就是

$$\underset{\sim}{A}=(0.35,0.2,0.2,0.15,0.1)。$$

于是，得评判结果为模糊集合（矩阵形式）$\underset{\sim}{B}=\underset{\sim}{A}\circ\underset{\sim}{R}$，即

$$\underset{\sim}{B}=(0.35,0.2,0.2,0.15,0.1)\circ\begin{pmatrix}0.2 & 0.5 & 0.3 & 0\\ 0.1 & 0.3 & 0.5 & 0.1\\ 0 & 0.4 & 0.5 & 0.1\\ 0 & 0.1 & 0.6 & 0.3\\ 0.5 & 0.3 & 0.2 & 0\end{pmatrix}$$

$$=((0.35\wedge 0.2)\vee(0.2\wedge 0.1)\vee(0.2\wedge 0)\vee(0.15\wedge 0)\vee(0.1\wedge 0.5),\\(0.35\wedge 0.5)\vee(0.2\wedge 0.3)\vee(0.2\wedge 0.4)\vee(0.15\wedge 0.1)\vee(0.1\wedge 0.3),\\(0.35\wedge 0.3)\vee(0.2\wedge 0.5)\vee(0.2\wedge 0.5)\vee(0.15\wedge 0.6)\vee(0.1\wedge 0.2),\\(0.35\wedge 0)\vee(0.2\wedge 0.1)\vee(0.2\wedge 0.1)\vee(0.15\wedge 0.3)\vee(0.1\wedge 0))\\=(0.2,0.35,0.3,0.2)。$$

但因 $0.2+0.35+0.3+0.2=1.05$ 不容易按百分数计算，可把评判结果改写为

$$\underset{\sim}{B}'=\left(\frac{0.2}{1.05},\frac{0.35}{1.05},\frac{0.3}{1.05},\frac{0.2}{1.05}\right)=(0.19,0.33,0.29,0.19)。$$

评判的结果表明，19%的专家认为这位教师的课堂教学效果优秀，33%的专家认为教学效果良好，还有29%的专家认为教学效果一般，而认为教学效果差的专家只有19%。根据模糊数学上的最大隶属原则，结论是这位教师的教学效果“良好”，因为最高隶属度和“良好”最为贴近。要比较几个教师的教学效果，也可以先按此法判定，然后再从评判结果按隶属度高低排其优劣次序。

以上整个评判过程，可以编成计算机程序在计算机上实现。

用模糊综合评判来评判教师课堂教学效果的好坏，由于运用了模糊数学的概念，对教学中带有“模糊性”现象的诸因素可建立起模糊关系，并利用模糊数学方法进行运算，其验证得到的结果比用一般的总评分或加权平均分，能更准确、科学地反映出该教师的教学水平。这不仅为评价教师教学效果提供了数量依据，而且还打破了以往那种定性的主观评价的方法。

不难看到，在对教学效果的评判中，权重的分配是一个至关重要的问题，应当尽可能使之符合实际。由此可见，衡量教师教学效果，既不能片面地听取学生的意见，又不能忽视他们的意见。如果大多数学生说“老师讲课我们听不懂”，那么无论如何也不能说这位教师的教学效果是良好的；即使大多数学生说“老师教得好，一听就懂”，也不能肯定这位教师的教学效果就是优秀的，因为还得判断教师讲授的深广度和在传授知识的同时对学生能力的培养如何，特别是在大学里，教师讲课一般很难使学生一听就懂，何况教学过程还不单纯是一个传授知识的过程。

四　模糊综合评判在高等教育评估中的应用

由于模糊评判是对人脑评判事物的思维方法进行模拟，因此它能对诸如教学效果“优秀”、“良好”、“一般”、“差”等评语作出适当的定量描述。但对高等教育进行评估时，不仅需要考虑的因素很多，而且因素之间还有不同的层次，这时如果再应用上述模糊综合评判（称为第一级模糊评判），权重就难以细致分配。

即使诸因素一一定出了权重，由于对权重 $\underset{\sim}{A}=\{a_1,a_2,\cdots,a_n\}$ 需要满足 $\sum_{i=1}^{n}a_i=1$，每个因素所得的权重 a_i 一般来说都很小，而模糊关系矩阵的合成运算是行和列的对应元素，先“取小”而后“取大”，较小的权值通过“∧”运算全被筛选掉了，所有单因素的评判失去了作用。在遇到像对高等教育评估这样多因素多层次的情况时，往往先把因素集合 X 按其某种属性分成 s 个子集：$X_1,X_2,\cdots,X_s$，且满足 $X_1\cup X_2\cup\cdots\cup X_s=\bigcup_{i=1}^{s}X_i=X$，$X_i\cap X_j=\varnothing(i\neq j)$，$\varnothing$ 为空集。设每个子集 $X_i=\{x_{i1},x_{i2},\cdots,x_{in_i}\}(i=1,2,3,\cdots,s)$，$\sum_{i=1}^{s}n_i=n$。再对于每一个 X_i 按第一级模糊评判的方法进行模糊评判。设 X_i 的单因素评判为 $\underset{\sim}{\boldsymbol{R}}_i$，于是得第一级模糊综合评判为

$$\underset{\sim}{\boldsymbol{B}}_i=\underset{\sim}{\boldsymbol{A}}_i\circ\underset{\sim}{\boldsymbol{R}}_i=(b_{i1},b_{i2},\cdots,b_{in})(i=1,2,\cdots,s)。$$

将每个 X_i 作为一个因素看待，用 $\underset{\sim}{B}_i$ 作为它的单因素评判，这样

$$\underset{\sim}{\boldsymbol{B}}=\begin{pmatrix}\underset{\sim}{\boldsymbol{B}}_1\\\underset{\sim}{\boldsymbol{B}}_2\\\underset{\sim}{\boldsymbol{B}}_s\end{pmatrix}=(b_{ij})_{s\times m},$$

即为 $\{X_1,X_2,\cdots,X_s\}$ 的单因素评判矩阵，每个 X_i 作为 X 中的一部分，反映了 X 的某种属性，可以按照它们的重要性给出权重分配，即

$$\underset{\sim}{\boldsymbol{A}}=(a_1^*,a_2^*,\cdots,a_s^*),$$

于是有第二级的模糊综合评判为

$$\underset{\sim}{\boldsymbol{B}}^*=\underset{\sim}{\boldsymbol{A}}\circ\underset{\sim}{\boldsymbol{B}}=\underset{\sim}{\boldsymbol{A}}\circ\begin{pmatrix}\underset{\sim}{\boldsymbol{A}}_1\circ\underset{\sim}{\boldsymbol{R}}_1\\\underset{\sim}{\boldsymbol{A}}_2\circ\underset{\sim}{\boldsymbol{R}}_2\\\cdots\\\underset{\sim}{\boldsymbol{A}}_s\circ\underset{\sim}{\boldsymbol{R}}_s\end{pmatrix}。$$

有了第二级模糊综合评判的数学模型，再模仿以上步骤，可得出更多级的模糊综合评判模型。□

计算机：诞生于想象力和数学

1 计算机并行算法和计算复杂性*

一　并行与并行算法

要使用计算机求解某个问题，必须首先根据问题的要求设计出一种算法，然后利用某一种算法语言编成程序并输入到计算机里。计算机的任务是执行程序，经过翻译解释，最后产生一个或多个输出，即得到要求的答案。

算法是计算机科学的核心课题之一，不论是从事计算机硬件设计还是从事软件设计，都必须认真研究算法。早在电子计算机出现以前，有人就已开始了算法的研究，并创立了许多有效的算法。如公元前两千多年，古希腊数学家欧几里得(Euclid)就创立了求两数的最大公因数的算法。我国战国时期的孙子创立了求若干数的最小公倍数的算法。举世公认的我国经典数学名著《九章算术》中所讲的解联立一次方程(线性方程组)的方法，就是其后18世纪德国数学王子高斯(Gauss)的解线性方程组的消元算法，这更是一种普遍有效的算法，至今仍被广泛应用。1946年电子计算机的问世与发展，使算法的研究取得了空前的进展。到了20世纪70年代，随着大规模集成电路的出现以及计算机的更新换代，算法的研究又发生了一次飞跃，产生了与并行处理机相适应的并行算法，从而使很多难于处理的问题得以迎刃而解。80年代中期国际上出现

* 本文原载于《自动化博览》1998年第3期。

了所谓“并行热”，各种并行计算机陆续出现，商品化的并行计算机大量投入市场。“并行热”的出现正是适应了当前科学技术发展的要求。

并行计算机是各类计算机中性能最好、功能最强的计算机，专门用来解决那些在一般巨型计算机上解决不了的科学计算、工程设计和数据处理等问题。现在，越来越多的人开始认识到并行化是科学计算和工程设计所必由之路。事实上，今天的并行计算机在诸多的科学技术和工程应用中起着关键的作用，如核武器设计、实时跟踪、气象预测、地质勘探、人工智能以及基础研究等等，都强烈要求快速而高效的计算机来实现大规模的科学计算和数据处理，以推动人类高科技发展。

首先，算法的优劣与计算机的结构密切相关。在某种计算机上运行得很好的算法，在另一类计算机上也许很差。传统的计算机是遵循数学家冯·诺伊曼(John von Neumann)的设计思想而设计并制造出来的，所以又叫做冯·诺伊曼计算机，它的特点是数据和程序一起存放于存储器，且按照预定的顺序一条一条地执行指令。这种按部就班地“串行”工作方式，显然不利于提高运算速度。因此，为了达到更高的处理性能，除了提高元器件的速度外，还必须改革计算机系统的结构。计算机系统结构的改进，主要是围绕在同一时间间隔内增加操作量，即所谓并行处理技术。为了并行处理而设计的计算机系统称为并行计算机(parallel computer)。在并行计算机上求解问题的过程称为并行计算(parallel computing)，在并行计算机上设计求解给定问题的算法称为并行算法(parallel algorithm)或称平行算法。一个算法并行性不好，就会从亿次机降为万次机。

按照上面的定义，从技术角度看，并行算法的研究无疑是并行处理的一部分，但它本身又有相对的独立性。由于当代电子技术和工艺的进步，从硬件实现计算机并行体系已不成为障碍，因此并行处理的主要困难在于算法和软件。

并行算法不只是一种新技术，而且也是一种新思想、新概念。J. Rice 曾指出，由于客观物理世界是并行的，这是人们开展并行算法的基础。正因为世界上很多事物是并行发生的，彼此之间又有一定联系，从而给并行算法的研究方法奠定了基础。

我们知道，算法是解题方法的精确描述，它是一组有穷的规则，这些规则确定了解决某一特定类型问题的一系列运算。所谓并行算法，简单地讲，就是适合于在各种并行计算机上求解问题的算法。1980 年，Kung 曾经给并行算法下了一个精确的定义：并行算法是一些可以同时执行的诸进程的集合，这些进程相互作用和协调动作，从而达到对给定问题的求解。

并行的含义是显而易见的，并行现象在日常生活中也屡见不鲜。譬如，许

多人排成一路纵队通过某一地段是串行，排成一路横队通过则是并行。在同一时间内，并行比串行通过的人较多。在电路中也有类似情况，并联电路比串联电路在相同的时间内可以通过较多的电荷。

并行算法可以从不同的角度进行分类：① 数值计算和非数值计算的。所谓数值计算是指基于代数关系的一类计算问题。诸如矩阵计算、多项式求值、解线性方程组等，它基本上属于数值分析（对以数字形式表示的问题求数值解）的范畴。所谓非数值计算，是指基于比较关系的一类计算问题，诸如排序、选择、搜索、匹配以及图论等方面的计算问题。它基本上属于符号（如字符、数字、图像等）处理的范畴。因此，我们把面向数值计算和非数值计算的并行算法分别称为数值计算和非数值计算的并行算法。② 同步的、异步的和分布式的算法。同步算法是指某些进程必须等待其他进程的一类并行算法。因为一个进程的执行，依赖于输入数据以及系统中断，所以全部进程均必须同步在一个给定的时钟，以等待最慢的进程。通常把运行在 SIMD 机模型上的并行算法称为同步并行算法。所谓异步算法，是指诸进程的执行一般不必相互等待的一类并行算法。在这种情况下，进程间的通信一般是通过动态地读、取（修改）共享存储器的全局变量来实现。通常把运行在 MIMD 共享存储机模型上的并行算法称为异步并行算法。所谓分布式算法，是指基于异步通信模型上设计的一类并行算法。这一类算法的典型特征是，通过消息通信来协同整个问题的求解。通常把运行在 MIMD—CL 机模型上的并行算法称为分布式并行算法。

二　算法的计算复杂性

研究算法的最终目的是为了有效地求出问题的解，这就需要将算法放到计算机上运行，因此对算法的效率要作出分析。分析算法不仅是一种有趣的智力活动，它可以在这方面充分发挥我们的聪明才智；更重要的是，从经济观点来看，分析算法可以使我们知道为完成一项任务所设计的算法的好坏，从而促使我们去设计一些更好的算法，以收到少花钱多办事办好事的经济效益。

求解同一个数学问题，可以有不同的算法。如果有多个算法都能求解同一个数学问题，如何对它们进行比较，以找出其求解的最有效算法？如何评价或衡量一个算法的优劣程度？这些问题无论在理论上还是在实践上都是十分重要的。这也就是所谓算法的计算复杂性问题。

算法的计算复杂性（computing complexity）也称复杂度，是衡量算法计算难度的尺度，人们可以从不同的角度来衡量和评价。使用最普遍的评价标准是

一个算法需要耗费的时间和空间(处理机台数或存储单元)。如果某一个算法所需要的时间或空间少,就可以说这个算法比另一个算法好。无论是从时间方面考虑还是从经济效益等方面考虑,人们总是希望使用效率高(消耗的时间或空间较少)的算法。时间和空间已成为评价算法的两个重要方面,也是算法设计中长期追求的目标。当我们设计一个算法的时候,不但要注意算法的可行性,更要关心算法与它能求解的问题规模 n 之间的关系,即随着问题的规模 n 越来越大时,该算法所需要的时间和空间究竟会怎样变化。问题的规模 n 视不同的问题而有不同的含义,如对矩阵来说是指阶数,对线性规划来说是指问题中的变量个数及约束条件个数等等。

如果一个计算问题的规模是 n,则把求解这个问题的某个算法看成是 n 函数,算法的时间复杂性(time complexity)或称为时间复杂度记为 $T(n)$。当 $T(n)$ 随 n 的变化而趋于无穷大时,我们就得到渐近时间复杂性。一般地说,我们关心的是 n 较大时的时间复杂性,这时它与渐近时间复杂性相差不多,所以对两者就不予区分。然而由于计算机的速度不同和指令系统不同,从而所用时间的多少随着计算机的不同而有很大差别。因此,在算法分析中,一般用初等运算(算术运算,比较和转移指令等)的步数表示一个算法在假设的计算机上执行所需要的时间,即每做一次初等运算,需要一个单位时间。类似地可以定义算法的空间复杂性(space complexity)和渐近空间复杂性。因为在大多计算问题中,所需的时间与空间相比,时间是主要的。因此,通常讨论较多的是时间复杂性,而很少涉及空间复杂性。

在对算法的时间复杂性进行分析时,常使用上界及下界概念。假使 $f(n)$ 和 $g(n)$ 是定义在自然数集合 $\mathbf{N}$ 上的两个函数,如果存在两个正常数 c 及 n_0,使得对所有的 $n \geqslant n_0$,均有 $f(n) \leqslant c \cdot g(n)$,则称 $g(n)$ 是 $f(n)$ 的一个上界(函数),记作 $f(n)=O(g(n))$。如果对所有的 $n \geqslant n_0$,均有 $f(n) \geqslant c \cdot g(n)$,则称 $g(n)$ 是 $f(n)$ 的一个下界(函数)。

对于某一正数 c,如果一个算法处理 n 个输入数据所需要的时间是 cn^2,则我们说算法的时间复杂性为 $O(n^2)$,并说时间复杂性是 n^2 量级的。大写字母"O"是描述算法性能的工具,它表示该问题的计算复杂性上界函数。算法复杂性量级是反映算法性能的最主要的指标。一个算法的复杂性量级越低,它的性能就越好,在计算机上运行的效率就越高。也许有人会认为,亿次机出现以后,算法效率无关紧要。然而,许多大型课题,如战略武器设计、气象预测、人造卫星上天等等,在亿次机上运行的时间往往长达数十小时,甚至上百小时。即使算法复杂性的一点小小的改进,也可获得节约资源、赢得时间、增加数据输入量的明显效果。为了具体说明这个问题,让我们来看以下例子。

假设某一规模为 n 的计算问题，有 5 个不同的算法 $A_1 \sim A_5$（如下表），它们的时间复杂性是加工 n 个输入（数据）所需要的单位时数。如果 1 单位时间为 1 毫秒时，那么在 1 秒钟里，算法 A_1 能处理完 1 000 个输入数据，算法 A_2 可处理完 140 个输入数据，而算法 A_5 至多只能加工 9 个输入数据。下表给出了这 5 个算法分别在 1 秒、1 分钟、1 小时内处理输入数据的上界。

算法	A_1	A_2	A_3	A_4	A_5
时间复杂性 $T(n)$	n	$n\log_2 n$	n^2	n^3	2^n
1 秒钟内能处理的输入量	1 000	140	31	10	9
1 分钟内能处理的输入量	6×10^4	4 893	244	39	15
1 小时内能处理的输入量	3.6×10^6	2×10^5	1 897	153	21

此表清楚地说明了各种算法由于计算时间复杂性不同，它们在同一时间内所能求解问题的规模相差极大。现在来比较一下同一计算机上使用不同算法的效果，以 1 分钟作为比较的基准。如用算法 $A_3=n^2$ 取代算法 $A_4=n^3$，则能求解的问题规模扩大 6 倍；若用 $A_2=n\log_2 n$ 取代 A_4，则能求解的问题规模扩大 125 倍；若用 A_2 取代 $A_5=2^n$，则能求解的问题规模扩大 326 倍。若以 1 小时作为比较基准，则差别就更为显著。所以我们说算法的时间复杂性是算法好坏的重要尺度，计算机速度越快，这一测度就越为重要。这是因为算法 $A_5=2^n$ 的时间复杂性函数是输入量 n 的指数函数，而算法 A_1,A_2,A_3,A_4 的渐近时间复杂性都是以多项式为界的。特别是算法 A_1，它的渐近时间复杂性是输入量 n 的线性函数。在这种情况下，随着计算机速度的提高，机器在同一时间里所能处理的输入量可以成倍地增长。计算机科学家普遍认为：如果一个算法的时间复杂性是以多项式为界的，则认为这个算法是一个有效算法。也就是说，对于一个有效算法，哪怕输入量很大，现代的计算机也还可以处理得了。如果一个算法的时间复杂性是输入量 n 的指数函数（且底数大于 1），则认为这是一个低效率的算法。对于一个时间复杂性是指数函数的算法，只要问题的输入量稍微大一些，现代计算机就处理不了，因为像 $A_5=2^n$ 这样的算法，当问题的规模 n 增大到一定程度后，所需时间急剧增加，任何高速计算机都难以满足。由此可见，算法的时间复杂性函数是衡量算法效率高低的重要尺度，不论是对串行算法还是对并行算法都是如此。

这样，从计算时间上可以把算法分成两类，凡可用多项式来对其计算时间限界的算法，称为多项式时间算法（polynomial time algorithm），有时也简称多项式算法；而计算时间用指数函数限界的算法称为指数时间算法（exponential

time algorithm)。例如，一个计算时间为$O(1)$的算法，它的基本运算执行的次数是固定的，因此，总的时间由一个常数(即零次多项式)来限界。而一个时间为$O(n^2)$的算法则由一个二次多项式来限界，计算时间为m阶的多项式的算法，其时间可用$O(n^m)$来表示。以下六种计算时间的多项式时间算法是最为常见的，其关系为：

$$O(1)<O(\log_2 n)<O(n)<O(n\log_2 n)<O(n^2)<O(n^3)。$$

指数时间算法一般有$O(2^n)$、$O(n!)$和$O(n^n)$这几种，其关系为$O(2^n)<O(n!)<O(n^n)$，其中最常见的是时间为$O(2^n)$的算法。当n取很大值时，指数时间算法和多项式时间算法在所需时间上非常悬殊，因为我们根本就找不到一个这样的m，使得2^n囿界于n^m。

一个问题的计算复杂性的下界，是这个问题本身所固有的。例如，用比较判别法从n个整数中找出最大的数，其计算复杂性下界为$(n-1)$次比较。如果一个算法的时间复杂性已达到了该问题的下界，算法不可能进一步改进。达到下界的算法称做最佳算法。但是，一个问题的固有计算难度或复杂性的下界是很难确定的。

但是，现实生产或生活中要计算的问题是纷繁复杂的，是否一切要求“计算”的问题都可以设计出算法呢？回答是“否”。按算法论和计算复杂性的观点，现有的问题大致可分为三类：

第一类是无算法的问题，如著名的“哥德巴赫猜想”、“费马猜想”均属于这一类。所谓猜想，是由人们的直观或直觉上的初步判断认为可能成立，而又未经严格证明的命题。通过严格的数学方法来论证，被证明了的猜想，就直接变成定理。所以一个好的、深刻的猜想，往往成为数学家们长期研究的课题。如地图的着色中，有所谓“四色猜想”，困扰了人们一百多年，终于在计算机的帮助下，成功地得到证明，使“四色猜想”变成了“四色定理”。

第二类是有多项式时间算法的问题，被称为P类问题(P是多项式polynomial的第一个字母)。这一类问题是大量存在的。例如最短路程问题(在一个网络图中求给定两点之间的最短路程)的迪克斯特拉算法是$O(n^2)$的，最大流问题(若把一个网络图看成是一个公路网，最大流问题就是通过这个公路网的最大运输量)的卡扎诺夫算法是$O(n^2)$的，这些都是著名的多项式时间算法。

第三类是有些问题虽能写出其算法，但被认为不可能存在多项式时间算法。这样的问题有整数规划(要求答案是整数的线性规划)和旅行推销员(货郎担)问题(一个推销员(货郎)要到n个村庄去推销商品，最后回到原出发点，求最短巡回路程)。但在20世纪70年代后期形成了一个新的理论，把上述两个问题及其他许多问题归成一类，称为NP—完备问题。直观地讲，一个NP—完

备问题是一个与任一个合理的问题同样困难的计算问题，且有下述性质：① 任何一个 NP—完备问题都不能用任何已知的多项式算法求解；② 若任何一个 NP—完备问题有多项式算法，则一切 NP—完备问题都有多项式算法。根据这两个事实，很多人猜想任何 NP—完备问题都没有多项式算法，但是无人能证明它。事实上，现在人们相信，不发展全新的数学技术就证明不了这个猜想。通常对这一类问题的研究或者是满足于寻找一些求近似解的方法，或者是考虑这些问题的特殊情况，因为一个 NP—完备问题的某些特殊情况可能是多项式时间的。当然 NP—完备问题没有多项式时间算法，目前还只是一个猜想，无论证明或推翻这个猜想，都将成为数学和计算机科学领域的一件大事，人们正翘首以待。□

2 数理逻辑与计算机*

逻辑学是研究关于人类思维的形式和规律的科学。人的思维及其语言表达都必须符合逻辑的规律，违反了便成为诡辩。

数理逻辑是用数学方法来研究逻辑思维的一门科学，是数学方法在形式逻辑领域内的应用。数理逻辑像现代数学一样，系统地使用符号、公式来陈述、处理自己的问题，对理论中的概念作出明确的定义，对定理作出严格的证明等等。总之，第一，使用特殊的符号将理论符号化；第二，由公理推出定理。所以，数理逻辑从它的研究对象来说，是逻辑学的一个分支。但是，由于它采用了数学方法，这就使得它本身也兼成为一门数学。

数理逻辑从它诞生的那一天起，就与控制技术有着不可分割的联系。自从亚里士多德逻辑建立之后，希腊人就试图用某种机械来代替人的思维活动。17 世纪，德国数学家兼哲学家莱布尼茨在前人的基础上，提出用全新的数学方法取代传统的逻辑学。他说："我将作出一种通用代数，在其中一切推理的正确性将化归于计算。"17 世纪，法国物理学家帕斯卡(Blaise Pasacl)制成了世界上第一台会演算加法的计算器。莱布尼茨又在此加法器的基础上研制成能进行全部四则运算的计算器。既然计算可以用机器来实现，那么人的推理活动是否也可以用机器来进行呢？这是逻辑学家和自动控制论专家所共同关心的问题。但是在莱布尼茨那个时代，由于生产力水平低下，这种用机器

---

* 本文原载于《自动化博览》1994 年第 4 期。

模拟人脑思维的愿望未能实现。后人沿着莱布尼茨所期望的方向继续研究，最终成功地创立了数理逻辑。因此，人们公认莱布尼茨是数理逻辑的创始人。

19 世纪，英国数学家布尔（George Boole）先后出版了《逻辑的数学分析》（1847）和《思维推理的探讨》（1854）两书，主张对旧的传统逻辑进行彻底改革。布尔认为，逻辑关系和某些数学运算相类似，代数系统可以有不同的解释，把解释推广到逻辑领域，就可以构成思维的演算。布尔终于成功地创立了当今有名的布尔代数。

所谓布尔代数，简单地说，就是研究只有两个不同的数即“1”和“0”的代数。这里的“0”和“1”不具有传统数学中的大小意义，但恰好是二进制中的数符。布尔的工作为数理逻辑奠定了基础。后来，罗素和怀特的数学名著《数学原理》的出版，标志着数理逻辑的成熟。20 世纪，人们沿着布尔的方向，发展了开关代数（又称逻辑代数），并在 1946 年研制出了改变工业社会面貌的第一台电子计算机，莱布尼茨梦寐以求的以“机器模拟人脑思维活动”的愿望终于成为现实。

现在电子计算机的电路虽然比较复杂，但基本上是由一些开关电路组成的。所谓开关电路，是指具有“接通”和“断开”两种状态的电路。输入、输出信号具有两种状态的电路就是一种开关电路。逻辑门电路、双稳态触发器也都是开关电路。开关电路中，“通”和“断”两种状态分别用逻辑代数中的“1”和“0”来表示。为使计算机能够进行逻辑判断和推理，还需采用计算机所能接受的语言来表示判断的“真”和“假”。因此，在计算机中一般采用二进制，“1”表示逻辑判断的“真”，“0”表示逻辑判断的“假”。

电子计算机是数理逻辑和电子学相结合的产物。实际上，无论是作为计算机雏形的图灵机，还是作为电子计算机的数学工具的布尔代数，进而言之，无论是作为计算机核心的算法，还是作为程序设计的语言等，无不涉及到数理逻辑的知识和理论。因此，数理逻辑是计算机科学的理论基础之一。

数理逻辑在发展初期被看成是“逻辑代数”，只是将数学方法（基本上是一些代数方法）应用于形式逻辑，在形式逻辑中引进变量和方程，用方程式和计算代替推理，把形式逻辑当作代数处理。数理逻辑发展到今天，一般分为 5 个分支：逻辑演算、证明论、公理集合论、递归论和模型论。其中，逻辑演算部分既包含古典二值外延系统、非古典的构造性逻辑、多值逻辑、无穷逻辑等，也包含内涵逻辑、模态逻辑、规范逻辑等并非因数学的需要而发展的内容。证明论主要是由希尔伯特创立的，他认为各数学分支都可以通过一阶逻辑来形式化。逻辑系统是一组公理和推理法则。从公理出发，利用推理所得到的命题就是定理，这个推导过程就是一个证明。证明论就是研究证明的理论。递归论起源于证明论所创建的算法化方法，它研究判定问题、不可解度问题、计算复杂度问题和

能行性问题等。其中,能行性理论在控制论中起着突出作用。一类问题所谓是能行可解的,就是说存在某种确定的方法和步骤来判定这类问题中的任一个命题的真或假。例如,我们要实现某控制过程的自动化,其首要研究的问题是这一过程的算法是否能行的问题——算法描述问题。事实上,大量的控制过程是不可能算法化的,比如不可能找到一种可以解决所有数学定理证明的算法。能行性问题的研究为我们揭示了哪一类问题有机械地解决的可能,哪一类是不可能解决的。只有能行可计算函数才能由计算机进行计算,从而才能由计算机来实现某些功能控制。模型论主要是数学模型,不言而喻,这是各学科领域的一个重要但又很棘手的问题。近年来,数理逻辑在基本理论和应用方面都有了长足进展,模糊逻辑、模糊推理、模糊信息和模糊控制等崭新的研究课题相继展开。

布尔代数是数理逻辑的基本组成部分。布尔代数中的"0"、"1"一般表示两种相互对立的稳定状态。在电子计算机中,如表示电位的"高"和"低",电脉冲的"有"和"无",开关的"开"和"关"等等。计算机中二进制数的运算实际上就是对0和1的运算,而对0和1的运算,也就是对一对"矛盾"的概念如"高"和"低、"有"和"无"等进行逻辑的判断和推理。

布尔代数中的三种运算规则是:逻辑加、逻辑乘、逻辑非。以上三种逻辑运算是最基本的逻辑运算,以之构成的"或"、"与"、"非"门电路是最基本的逻辑电路。这三种逻辑电路都只有一种逻辑功能,叫做简单的逻辑门电路。计算机广泛采用的是由两个或三个简单门电路所组成的复合门电路。

随着人类大量思维过程的机械化、计算化的发展,逻辑代数的应用也在发展。除了在计算机线路设计方面继续发挥重要的作用外,它还是研究计算机软件的基础知识和重要工具。□

3 电脑与人脑

1946年世界上第一台电子数字积分计算机ENIAC的问世，是人类生活中的一件大事，它标志着科学技术发展史上的一个新的里程碑。大量的实践表明，电子计算机是人类脑力劳动的工具，它可以模拟、延伸人脑的功能。因此，电子计算机又被誉为电脑。

电子计算机诞生之后，在短短的40多年中，从元器件和软件方面看，它已经经历了以电子管和机器语言为标志的第一代，以晶体管和高级程序语言为标志的第二代，以集成电路和操作系统为标志的第三代，以大规模和超大规模集成电路和软件工程、网络软件为标志的第四代计算机的变革。但从整机原理和设计思想来说，并无实质性的突破，它们的体系结构始终没有跳出冯·诺伊曼型计算机的结构体系。现在的各种计算机系统都是按照冯·诺伊曼的基本结构设计的，所以都叫做冯·诺伊曼型计算机。具体表现在以下几个方面：

1. 都是由运算器、控制器、存储器、输入设备和输出设备这五大部分构成的。其中，运算器和控制器组成了中央处理器，负责运算与控制；存储器负责存储程序与数据。显然，存储与运算是分别集中进行的，且两者是分离的。

2. 工作方式是顺序的、串行的，因此它们所执行的算法都是图灵模型意义下的算法，即串行算法。这种工作方式易于形式化、数字化，这是计算机能够达到高精度的基本条件。另外，当人们需要用计算机求解问题时，必须事先用某种程序设计语言编制程序，具体地指出先做什么、后做什么以及怎样做，计算机只能按照预定的顺序一条一

条地执行指令，只能完成程序所指定的工作。

3. 所采用的元件只有两种状态，任何数据或符号在计算机内部都是用二进制数表示的。因此，需要用计算机求解的问题都必须事先把它转化为一系列的布尔代数运算，这样才能在计算机上实现。

4. 从功能上看，它们主要用于数值型数据的计算及非数值型数据的符号处理。

5. 人—机接口不自然，而且输入、输出的低速度与主机的高速度处理能力不匹配，使得人们不能更方便地使用计算机，它的效率亦得不到充分发挥。

现有传统的冯·诺伊曼型计算机实际上是模仿人的左大脑而设计的，其基本功能是算术运算和逻辑判断能力以及记忆能力。与人脑相比，其数值计算的速度单机已可高达每秒数十亿次（一次代表做一次加法），这是人脑望尘莫及的；其计算精确度高则是它的另一个特点；此外，通过扩展外存，它的记忆力几乎是无限的。应该说在这三个方面计算机确实大大地扩展和延伸了人的大脑功能。尽管计算机可以发挥速度和精度优势，完成在时间上人不可能做到的事情，然而功能再强、性能再好的计算机依然不可能替代人的大脑。究竟计算机和人脑在工作方式和特点上有何不同？用前四代冯·诺伊曼型计算机与人脑相比，可明显地看出它有如下一些局限性：

1. 人的记忆与思维是相随相伴而不可分的，信息是分布式存储的，但冯·诺伊曼型计算机的数据处理与存储是完全分离的。在处理器与存储器之间仅仅通过一条狭窄的通道逐字地交换数据，这就与大脑中记忆与思维合一而且信息分布的方式不一致，不能满足模拟人类记忆与思维的需要。

2. 人的思维过程是串行与并行共存且以并行为主的，这就使得人们不仅可以顺序地处理问题，而且可以同时应付多种不同的场景，把问题的不同侧面、不同因素密切地联系起来，进行多方位的综合性的思考。但冯·诺伊曼型计算机只能进行串行式的处理。

3. 现实世界中的事物并不都是“不为真就一定为伪”的两态逻辑，更多的是多态逻辑、非确定性的、模糊的。人脑既可进行精度不高的精确思维，更善于模糊思维。它可以用定性的、不精确的、模糊的术语进行亦此亦彼的模糊思维活动，对模糊事件进行推理处理，因而人脑可在低精确度条件下完成非常复杂的任务，而且可以达到相当高的可靠性。但冯·诺伊曼型计算机是基于两态逻辑的，为了对不确定性进行处理，需要从软件上弥补这一缺陷，这就又增加了软件设计的复杂性和难度。

4. 人的思维方式除了逻辑思维以外，还有形象思维、顿悟思维等，人们在处理问题时，通常是把它们结合起来进行的。但是，冯·诺伊曼型计算机只能进行符号处理，任何要在计算机上进行处理的问题都必须表示为一串符号序列，

并且还要给出处理这些符号的规则。因此它所解决的问题仅仅局限于逻辑思维所能解决的问题范畴内，对于那些需要形象思维的问题只能望尘莫及。这就使得冯·诺伊曼型计算机难以有创造性，它只能做人们为它规定的工作。冯·诺伊曼型计算机在思维能力方面的局限性极大地限制了它的功能，使得它无法进行音乐、美术方面的创作，数学上难以发现一个尚未发现的新定理。目前人工智能在自然信息理解、机器翻译、机器学习等方面遇到的困难都与冯·诺伊曼型计算机的这种局限性密切相关。

5. 人脑具有感知能力，能通过感觉器官感知外部世界，得到所需要的有关信息。视觉和听觉是人接受外界信息的主要通道。实验表明，人有80%以上的信息来自视觉，有10%以上的信息来自听觉，但冯·诺伊曼型计算机不具有感知能力，它需要人们通过输入设备把已表示出来的信息交给它进行处理。尽管人们已经通过软件的方法设计了各种比较方便的人—机接口方式，但它仍然是非自然性的，人们不能直接通过“说”或者手写的文字以及图像等来与计算机交互。此外，人们具有多种形式的表达能力及行为能力，能对外部刺激及时地做出反应，这也是冯·诺曼型计算机做不到的。

总之，虽然目前的冯·诺伊曼型计算机已经为人类做出了巨大的贡献，但它与人类的智能相比还有很大的差距。为了实现人类智能在计算机上的模拟、延伸、扩展，必须对其体系结构、工作方式、处理能力、接口方式等进行彻底的变革。正因为如此，20世纪70年代初期，国际上一门崭新的学科——并行算法开始形成并迅速发展起来。所谓算法，就是解题方法的精确描述，它包括一组有穷的规则，这些规则规定了解决某一特殊类型问题的一系列运算。并行算法是适合于各种并行计算机上求解问题和处理数据的方法，其形式定义是：并行算法是一些可用同时执行的诸进程之集合，这些进程相互作用和协调工作，实现对给定问题的求解。

并行计算机是使用计算机协同工作的一种高性能的计算机系统。其基本原理和特点是由操作数的到达来驱动指令执行，将数据流和指令流结合在一起，可进行全面的并行处理，并可顺利地提高运算速度，具有极高速化的特点。90年代以来，并行计算机一直是世界各国计算机研究的热点，代表着一个国家计算机研制水平的高低，也是解决大范围天气预报、地震分析、航空气动力学等大型计算问题的唯一有效途径。60年代中期，美国就开始了并行计算机的研究工作。70年代陆续研制出若干通用或专用并行计算机，并产生了与并行计算机相适应的并行算法。80年代初期，日本新一代计算机技术研究所在渊一博教授的带领下，开始研制第五代计算机。这种新一代计算机系非冯·诺伊曼型体系结构，它由知识库、解题与推理以及人工智能接口三大子系统构成。文件中知

识的更新以及对相关知识的检索，是由知识库管理软件执行。通过知识信息处理系统，知识可作为推理的基础。它采用分解法进行推理，并作为逻辑处理硬件的目标。Prolog 语言则是逻辑处理软、硬件之间相互通信的语言。第五代计算机具有自然语言理解能力以及图像处理能力，这不仅需要有能判定语言和图像最基本特征的硬件，而且还需要能有效使用知识库，以及理解语言和图像意义的软件（这种非冯·诺伊曼体系结构的新一代计算机原计划在 90 年代初期问世）。1992 年，中国国防科技大学研制出我国第一台通用 10 亿次并行巨型机——“银河—Ⅱ计算机”，可进行每秒 10 亿次以上的运算操作。1995 年“曙光 1000”大规模并行计算机问世，其速度高达每秒 25 亿次，这是我国在并行处理方面又一新的里程碑，被电子工业部评为 1995 年十大科技成果之一。1996 年，中国航空技术研究所的科研人员经过五年攻关，成功地研制成 PARGS 可伸缩并行处理计算机系统，这是继“银河—Ⅱ计算机”、“曙光 1000”计算机之后，我国自行研制的又一大型并行计算机系统，其运算速度可高达每秒 32 亿次，创造了我国自行研制计算机运算速度最高峰的记录。专家组鉴定认为，达到 20 世纪 90 年代国际先进水平。PARGS 的研制成功，标志着我国计算机技术又跨上一个新台阶。

毛泽东曾说过：“许多自然科学理论之所以被称为真理，不但在于自然科学家们创立这些学说的时候，而且在为尔后的科学实践所证明的时候。”计算机的发展正是经历了实践—理论—实践的过程，并且仍然在这个过程中不断地向前发展。最近几年来，计算机科学界出现了新的技术革命，就是在理论上与实践上探求智能计算机系统。所谓智能计算机，首先表现在“智能”二字上。智能的含义是，如果一个系统能够回答某个领域的问题集，并能解决问题，则称该系统具有这个领域的智能。因此，“智能”并不是无限的，要受领域的影响与限制。有人用一个简单的公式来表示：智能＝知识＋推理。这个公式刻画了智能计算机要体现人的思维活动。知识在推理的作用下将产生新的知识，因此智能计算机的一个突出表现就在于它可以产生知识，代替人脑的作用。□

4 非冯·诺依曼型计算机*

目前世界上运行的计算，尽管种类繁多，但按其数据加工方式可以分为两大类：串行计算机与并行计算机。

所谓串行计算机也就是冯·诺依曼型计算机。就其整机原理和设计思想来说，现在的计算机是根据美国数学家冯·诺依曼提出的原理设计并制造出来的。例如，1946年世界上出现了第一台电子数字积分计算机 ENIAC。ENIAC 的操作对象是以二进制代码形式表示的数据和指令，把它们一起存放在同一存储器中，同时使用了少量的寄存器用于存放当前执行的指令和数据，计算机不断地按顺序完成存取和执行指令的重复工作。从 1946 年起，冯·诺依曼的这种机器被广泛地接受，并适用于所有的计算机。

冯·诺依曼型计算机的一个基本思想是将数据和指令存放在一起，这一思想对于计算机的发展有着深远的影响。由于允许数据和指令随意地存放在同一存储器中，这就意味着从存储器中存取指令和存取数据的机制是完全一样的，而且也意味着一个符号可以在一个地方作为数据处理，而在另一个地方又可作为指令执行。现在使用中的大多数的计算机都是按冯·诺依曼的基本结构设计的，所以叫做冯·诺依曼型计算机。

但是随着科学技术的发展，采用多处理结构和多向量阵列结构的各种计算机相继问世，出现了许多通用和专用性质的大规模并行计算机，从而使计算机技术跨入了一个

---

* 本文原载于《电脑爱好者》1994 年第 6 期。

崭新阶段，商品化的并行计算机大量涌现，国际上出现了所谓“并行热”。并行计算机应用范围越来越广，许多问题诸如核武器设计、气象预测、实时跟踪、卫星图像识别、数据管理等都强烈要求快速而高效的并行计算机来实现大规模的科学计算和数据处理。由于当代电子技术和工艺的进步，从硬件上实现并行计算机体系已不构成障碍，因此，并行处理的困难主要在于算法和软件上。

什么是并行处理和并行算法？为什么并行算法比串行算法快？并行处理从广义上讲，并非是一种新思想，它早用于通用计算机系统设计中的多处理技术及多道程序技术。然而，两者的根本区别在于，在通用计算机系统中，系统设计的目的是为了获得一个对于混合作业最大的生产率（并行完成独立的作业），而这里的“并行处理系统”，其系统设计的目的是为了利用多个计算机资源去完成一个单一的作业（并行完成相关的作业段），以获得极高的处理速度。这样的例子在日常生活中也屡见不鲜。例如部队的“阅兵典礼”，许多士兵通过检阅台前受阅。如果排成一路纵队“串行”通过，则需要较长的时间，即速度较慢；如果排成一路横队“并行”通过，则需要的时间较短，也就是速度较快。在电路中也有类似情况，并联支路在相同的时间比串联支路通过较多的电荷。计算机设计者把这种道理用在计算机设计上，设计出了一种新颖的计算机——并行计算机，其主要特点是对数据做批量处理，即进行数组或向量运算。这与串行计算机很不相同，串行计算机只能做标量运算，即对单个或一对数据进行操作，所以并行计算机比串行计算机计算的速度要快。传统的串行算法往往很不适合在并行计算上运行，所以必须加以改造，以提高算法的并行性。如果一个算法的并行性不好，就会使并行计算机的效率大幅度下降，从亿次机降为万次机。正因为如此，20 世纪 70 年代初期，国际上一门崭新的学科——并行算法开始形成并迅速发展起来了。通俗地说，在并行计算机上求解问题和数据处理的算法称为并行算法。我们知道，算法是解题方法的精确描述，它是一组有穷规则，这些规则确定了解某一特定类型问题的一系列运算，在执行有穷步的运算后终止，并产生一个或多个输出。根据这一思想，1980 年，Kung 曾经给并行算法下了一个精确的定义：并行算法是一些可同时执行的诸进程的集合，这些进程相互作用和协调动作，从而对给定的问题求解。

并行计算机是可以同时执行许多类操作的计算机，是各类计算机中性能最好、功能最强的计算机，它具有巨大的数值计算和数据处理能力，专门用来解决那些在一般大型计算机上解决不了的科学计算、工程设计或数据处理等问题。核武器设计、气象预测、地质勘探、实时跟踪、图像识别、数据管理等都需要高速度、大容量的计算机，串行计算机系统的改进所获得的速度的提高已远远赶不上实际需要。20 世纪 60 年代中期，美国就开始了并行计算机的研制工作，70

年代陆续研制出若干通用或专用并行计算机，产生了与并行计算机相应的并行算法。日本政府也急起直追，投资4亿美元，进行第五代（新一代）计算机研制，现已开发出数据并行计算机软件，包括程序设计语言和并行处理机系统（据1992年7月7日《中国科学报》报道）。1992年，中国国防科技大学计算机研究所研制出我国第一台通用10亿次并行巨型机——“银河—Ⅱ计算机”，可进行每秒10亿次以上的运算操作，填补了我国通用并行巨型机的空白，再次缩小了我国与国外先进水平的差距，并使我国跨进了世界研制巨型计算机的行列。国家专家鉴定委员会认为，具有联网能力的“银河—Ⅱ”巨型计算机，既适合于作为超级计算中心的主机，又适合于作为大规模数据处理的中心机，有广泛的应用。

电子计算机的诞生是科学发展史上的一个里程碑。它发展到今天，从元器件方面看，经历了电子管、晶体管、集成电路的发展阶段，现在正向大规模集成电路阶段过渡；从使用方面来看，软件从无到有，现在已出现了工作量以几千人来计算的巨大软件工程，软件的价值正日益被人们所认识和重视。今天的计算机已拥有各种类型的操作系统，几十种计算机语言。计算机开创了人类脑力劳动的自动化，丰富了人类的精神财富，提高了人类对自然界的认识能力。尽管冯·诺依曼型计算机还是计算机发展的基础和主流，然而，由于冯·诺依曼型计算机的程序顺序执行方式严重地影响到计算机效率的提高，在冯·诺依曼模型基础上发展起来的软件系统已越来越复杂、庞大，其正确性无法保证，程序可靠性低下，加之软件生产率等问题构成了所谓的软件危机。软件代价不可控等问题使得计算机系统的整个性能大幅度下降。因此，计算机体系结构的设计者们不断地寻求各种途径，使计算机脱胎换骨，彻底变革，更新计算机结构，从根本上摆脱这种传统式的计算模型，以至于尽可能地开发并行系统，提高计算机的处理效能。于是，一种全新概念的非冯·诺依曼型计算机便应运而生。

非冯·诺依曼型计算机是基于与传统的控制流机器不同的计算机，其体系结构是属于非冯·诺依曼型的，如数据流计算机、归约机、动作体计算机、人工智能计算机和函数计算机等等。

所谓数据流计算机是一种从数据流的概念出发，由操作数的全部到达作为标志来驱动而不是由指令来驱动程序执行的机器，因此当多个操作同时满足这一条件时，它们可以并行地执行而不受程序的顺序限制。它基于两个原则：一是异步，对任何一条指令的点火执行当且仅所有操作数是可用的；二是函数性，所有操作都是函数，没有副作用。通常，数据流程序用数据流图来描述，数值作为标记通过数据流图。

数据流计算机模型支持的是函数语言，它的全部操作都是函数。点火规则

的制定能充分开拓并行性，而且这种语言所书写的程序易读易维护，软件生产率可以提高，显示出强大的生命力。数据流计算机具有并行性，加之以上的优点，近年来已引起人们广泛的注意，但由于数据流非常依赖于高技术和应用领域，因此需要我们进行更多的研究和开发工作。

归约机的发展要比数据流机来得晚些，它是要求驱动模式，一条指令的激发执行当且仅当有其他指令要求该指令的结果。归约程序由嵌套表达式组成，在归约中最接近指令的是函数应用。更高层的程序结构，指令和参数都是表达式。串归约和图归约是两种基本的归约程序组织。串归约实现的是“按值调用”的数据机构，而图规约则为“按引调用”的数据机构。然而这两种形式的归约都具有“递归”控制机构。

动作体计算机是面向目标的计算机，它采用面向目标的程序设计语言，这是一种高级的描述性语言，具有良好的模块化程序设计能力。在这种语言中，所有操作对象都叫目标。目标即执行的数据，是计算的基本单位，它实质上是一个信息包和对它操作的描述，因此目标一般具有内部状态，当它接受信息后就变为活动态，并执行所规定的操作。

函数计算机，也即 LISP 机，是一种面向 LISP 语言的机器。该机器设计成适合于 LISP 语言执行的机构，即所谓栈结构，具有速度快，低开销的优点。LISP 语言是一种计算机表处理语言，它没有语句的概念，整个程序由一组函数组成，广泛应用于人工智能领域。

最后简单谈一下人工智能计算机。这是一种有知识、会学习、能推理、体现人类思维活动的计算机，是新一代计算机的一种类型。例如日本近年来研制的第五代计算机就属于智能计算机。它由若干个子系统构成，包括从小到大专门支持某种应用的各种类型的机器和相应的软件。这些机器连接起来组成一个分布系统，系统中主要的部分(子系统)是推理机、知识库及知识库机。如果没有这两个子系统，就称不上为智能机。

以上所谈的数据流计算机、归约机、动作体计算机、函数计算机、智能计算机等，由于突破了冯·诺依曼设计思想的框架，为了便于区别，人们通常称它们为非冯·诺依曼型计算机。□

5 漫谈计算机软件*

我们知道，计算机系统是由硬件及软件两部分组成的。由主体及外围设备构成的实体，称为计算机的硬件或硬件设备，又叫硬件系统。主体包括存储器以及组成中央处理机的运算器与控制器，外围设备就是输入输出设备。但这仅是计算机的物质基础。现代计算机硬件虽然提供了很强的功能，例如在完成复杂数字运算方面远在人脑之上——一台计算机每秒钟可以进行几百万次乃至上亿次的运算，另外还具有逻辑判断功能，但对使用者来说，却并不总是适用和有效的。如现在计算机的价格是很昂贵的，因此不应只让一个用户独占一台计算机，应允许几个用户同时共享，使其充分发挥作用。为此，计算机就需要有一个管理和控制部件，负责用户之间的分配和组织其运行。但是，这种功能并不是硬件本身所能实现的，而需由计算机软件来和它配合完成。一台计算机如果只有硬件而没有软件，犹如一个人光有骨骼而没有肌肉与灵魂，是干不了任何工作的。在硬件的基础上，对硬件的功能实现扩充与完善，使计算机更加适用、更加经济，这就是计算机软件的任务。所以计算机是由硬件与软件不可分割的两个部分组成的，只有这两个部分有机地结合在一起，才能充分发挥计算机的作用。

那么，什么是计算机软件呢？软件是相对于硬件而言，是泛指使用计算机时所必需的各种各样的程序，或者说是程序集合。因此，软件又可称为程序系统，也称软件

---

*本文原载于《自动化博览》1992年第3期。

设备。程序是软件的核心，也是计算机科学上的一个重要的基本概念。什么是程序？顾名思义，程序就是“次序”的意义。计算机工作时得有个次序，先做什么，再做什么，这就是计算机程序。稍严密些，计算机程序就是根据计算机的指令系统，对于要解决的问题安排的计算步骤。它是由各条指令组成的（指令就是指明机器进行一种什么操作的命令）。计算机依次执行程序中的每条指令，以获得它所期望的结果。但设计某一程序的目的不是为了解决某一具体的计算或数据处理问题，而是为了完善和扩充计算机的功能，使硬件与软件结合成一台更为理想的计算机，从而有助于用户程序的编制、调试和运行。因此，从一般用户的角度看，这些软件和计算机中的中央处理机、存储器等硬件同样是计算机中不可或缺的部件，只不过硬件是由电子元件组成，而软件是由程序组成的。

习惯上，人们把计算机软件分成系统软件和应用软件两部分。

一　系统软件

系统软件不是为了解决某个特定的问题所编制的程序，而是指用于指挥整个计算机系统运行的操作系统。它着眼于利用计算机本身的逻辑功能，合理地组织整个解题过程，简化或代替在各个环节中由人所承担的工作，从而达到充分发挥和提高机器的效能，便于用户掌握使用计算机的目的。因此，它可用于计算机的管理、维护、控制、运行以及计算机程序的翻译、管理、维护、控制和运行。归纳起来，系统软件本身又可分成操作系统、语言处理系统和常用服务例行程序。

1. 操作系统。这是计算机系统中最重要、最基本的软件之一。它在硬件的基础上实现了硬件的第一次扩充，或者直接为用户使用计算机系统提供了可能与方便。正是因为操作系统有很大的优越性，所以它已被广泛研制与使用。从大型计算机系统到微型计算机，一般都配备操作系统。操作系统大致有如下几种：① 批量处理系统（把作业成批地交给计算机操作系统控制执行的方式称为批量处理，相应的操作系统称为批量处理系统）；② 分时操作系统（如果多个用户可以借助于一个操作系统，通过各自的终端共享一个计算机系统，这种作业处理方式称为分时处理，相应的操作系统即称为分时操作系统）；③ 实时操作系统（每当有一次处理要求或每当产生一次数据，要求计算机必须在规定的时间内立即进行处理并作出响应，这种处理方式称为实时处理，相应的操作系统即称为实时操作系统）；④ 网络操作系统（用通信线把分布在不同地点的多台计算

机连接起来所构成的网络称为计算机网络，组成网络的计算机有其自己的操作系统，为把网络中所有计算机有机地连接起来而设计的操作系统称为网络操作系统）。

操作系统本身不是解题程序，并不直接完成计算任务，它的作用是管理控制系统资源。所谓资源即凡作业和进程所要使用到的或进程中所要依赖的一组对象，都可以是该作业或进程所要求的资源。系统提供多种不同类型的资源。有些类型资源是由硬件直接提供的（如处理机、存储器、外部设备等），有些类型资源是由软件提供的（如文件、过程、信息等）。由于资源众多，操作系统的规模日益庞大与复杂，因此必然要带来一些新的问题，如可靠性（一个系统在使用多年之后，可能要出现一些错误，即所谓硬件的故障），开销和维护费用（操作系统自身耗费的资源称为开销，操作系统越大，开销也就越大；另外，由于操作系统的规模较大，在交付使用时难免仍保留有相当数量的隐患，造成日后的软件故障。因此，除了硬件维护外，还需要软件维护，软件维护费用过高，已成为一个突出问题），还有研制周期等。由于这些问题的存在，推动了操作系统自身的研究，也促进了计算机科学的发展。出于在尽可能短的研制周期内获得可靠的大型软件的动机，程序设计方法学、程序正确性证明、软件量度学与软件工程学等等的研究得到了深入的探讨，并已获得了相当大的进展。例如，一门研究软件开发与维护的普遍原理和技术的工程学科，即软件工程学从 20 世纪 70 年代初期开始迅速发展起来，现在已成为计算机科学的一个重要分支，一个异常活跃的研究领域。

2. 语言处理系统。机器语言就是计算机的指令系统，语句实际上就是指令。其特点是，不经任何翻译，计算机直接就能理解它。计算机按照编排好的指令序列依次进行各种操作。这种指令序列，通常又称为程序。编排程序的过程叫做程序设计。每种计算机都有自己的指令系统。不论什么计算机，指令都是由操作码和地址码两个部分组成的。操作码是表示什么操作的代码；地址码表示参加操作的数据（或原始数据或运算过程的中间结果）存放在存储器（或寄存器）的哪一个单元内，或者表示要送到何处去。表示操作码的位数和操作种类的多少是密切相关的，表示地址码的形式更是多种多样。有的机器指令的操作码只和一个地址码有关，这样的指令叫一地址码；有的机器指令的操作码和两个地址码有关，叫二地址码；余可类推。

3. 常用服务例行程序（简称例程）。它是一些绝大多数用户都要使用到的、完成大多数用户所要求完成的任务的程序。它们是按照一定的标准例行规则编制的。常用服务例程的数目很多。如外部介质的转换例程（其任务是把信息从一种外部存储介质复制到另一种外部存储介质上去）、库存管理例程、编辑例

程(可使用户通过很简单的命令建立、生成和修改文件)、联结装配例程、医疗诊断例程等等。

二 应用软件

应用软件指的是用户为方便某种应用或为解决某类特定问题的求解所需的各种具体程序。这类特定的问题可以是科学计算的、数据处理的,也可以是实时控制的等等。凡是这样的应用程序统称为应用软件。例如用于经济计算、情报检索、档案管理、医疗监护、工程设计、宇航控制、交通指挥、博弈等领域的专用软件。据统计,目前这方面的软件已达几千种。软件的出现,使计算机不仅有效地模拟和代替人脑的某些功能,扩大了计算机的功能和用途,而且也大大方便了操作人员和用户。可以说软件是计算机的灵魂。

由于近二十多年来,计算机应用日益普及和深化,应用领域急剧扩大,从传统的科学计算和商业数据处理扩展到几乎所有的行业。硬件的发展和应用范围的扩大,促使计算机软件也大大向前发展。首先是涌现了大量的高级程序设计语言和许多专门用于书写软件系统的语言,从而为软件的生产提供了有力的工具,于是产生了种类繁多的应用软件系统。如应用软件包、软件库和面向问题的程序语言等等。

应用软件包是一个为实现某种特殊功能或特殊计算的程序组,而这些特殊功能或特殊计算是为许多用户所需要的。软件包中的程序可以很容易地从一台计算机搬到另一台计算机上去应用,也就是说,它们都是与机器无关的。软件包的种类很多,大凡应用计算机的行业都有适合本行业的应用软件包。

最后谈一谈硬件与软件的关系。计算机科学家们认为,硬件与软件是相辅相成的。硬件是软件发展的物质基础,软件要迅速发展,对硬件的稳定可靠又必须有严格的要求,因而也促进了硬件的发展。近年来还出现了一种具有软件功能的硬件,这是硬件与软件相互结合的产物,称之为"固件"(或稳定固件)。像含有微程序集的控制存储器,它能把某些软件的功能并到硬件中去,即所谓"软件硬化"。把硬件和软件融合起来,可使计算机的速度和效能得到改进。□

6 计算机解题*

1946年电子数字积分计算机(ENIAC)的出现，是人类历史上具有划时代意义的一件大事，它不仅把人类从繁重的体力劳动中解放出来，而且也代替了人的部分脑力劳动。随着计算机的更新换代，已逐步形成了一个比较完善的"软硬兼施"的计算机系统。它由硬件和软件两部分构成：硬件即通常说的计算机主体和外围设备，即系统的机器部分；软件是为了用户使用计算机方便和提高计算机的使用率而提供的各种算法语言、操作系统、应用程序等等，即系统的程序部分。硬件和软件一起构成了一个不可分割的整体——计算机系统。

作为高度自动化信息处理的设备——计算机，被广泛应用于对各种实际问题的求解。控制论创始人维纳(Wiener)曾说过："计算机本质上是一种记录数字、运算数字并给出数字结果的机器。"把一个实际问题在计算机上求解，通常称为计算机解题或计算机求解。

计算机解题的必要性。由于火箭、导弹、原子弹以及系统工程等高新科学技术的发展，需要求解许多极其繁复的数学计算课题。这些课题需做几千、几万乃至上百万或百亿次以上的运算，不可能由手工计算求解，这就迫使人们不得不借助于计算机求解。例如天气预报，现在已经可以把不同范围的气象资料排成许多方程，从而构成天气预测的数学模型。用大型高速计算机很快便可以求出该数学模型的解，可以预报未来几天里的天气。但如果用手工

* 本文原载于《自动化博览》1997年第3期。

来进行，必须大量简化数学模型，即使动员若干人同时工作，也难以得到精确结果，加之计算很慢，等到解出来，业已时过境迁，成了历史资料。天气预报时效性很强，必须运用计算机进行计算，从而及时做出准确的预报。现在半球、全球的大范围天气预报已提到议事日程上来，这些问题的解决，一方面要加速发展空间科学技术，另一方面也迫切要求使用高速度运行、大容量存储的计算机。只有这样，人们才能进入控制天气的新时代。

计算机解题步骤。每当给定一个数学问题，在交予计算机求解之前还必须有一番数学加工和上机准备，而在计算机求解之后，又必须作计算结果的整理，检查是否符合实际问题需要。一般来说，计算机解题过程是：

1. 提出问题，建立问题的数学模型。所谓提出问题，就是提出要计算机解决的问题(例如天气预报问题)。但是每当给定一个实际问题，计算机并不能立即求解，因为它不能直接理解和接受所要求解决的实际问题。计算机所能理解和接受的只是按照实际问题所编制的程序，而程序是根据数学模型编制的。因此，当提出实际问题在计算机上求解时，首先是要建立该问题的数学模型。

数学模型是问题(或系统)的某种特征的本质数学表达式，即用数学式子(如代数方程、微分方程、差分方程等)来描述所研究问题的客观对象，或者问题在某一方面的存在规律。如导弹、火箭的运动轨迹的数学模型都可以归结为非线性微分方程。

一般来说，尽管所提实际问题的数学模型有解析表达式，即无穷多个连续自变量与函数值间的精确关系，但在计算机上求解时只能是离散的解，也即一组不连续的自变量与相应函数值近似地表达的函数关系，这表明计算机求解总是通过一个近似算法来求得近似解。

2. 确定算法，编制程序。为编制程序，需要把实际问题的解题方法确切地表示成一系列的可计算的步骤，其中每一个步骤的计算都是可由计算机执行算术运算或逻辑判断等操作来实现的，然后按照所确定的步骤来编成程序，提交计算机执行。可见，重要的是确定这一系列的计算步骤。只要所得计算结果符合问题的陈述，具有一定的代表性，在误差允许范围内，就承认所要求的解答为该问题的解。

当算法确定后，程序编制人员根据算法设计计算的先后次序，让每一个计算步骤用一组指令来实现。所有这些指令的全体构成一个程序。

3. 上机调试。在终端上编好程序后，启动主机，计算机就执行这个程序，计算结果由输出设备输出。

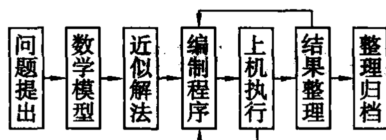
一般来说，对于稍大的程序难免存在错误。譬如，某几条指令编错了，导致与算法不一致；或者某数据写错了，致使结果不正确等等。每当查出错误，就得

对程序进行修改，然后再次上机。这个反复上机、反复发现与修正错误的过程，称之为程序的调试。

一个程序不能得到符合要求的结果，不仅可能是程序错，甚至可能是所确定的算法根本不正确，因此就得从确定近似解法、确定算法开始，重新编制程序，有时甚至应对问题的数学模型进行重新审核。

4. 整理归纳。对所得计算结果应该进行分析整理，检查是否符合实际问题的要求。当感到满意时，表明所选择的算法与所编程序都是正确的，无需再次上机。这时应总结全过程并整理成文件归档保管。

这样，计算机解题的全过程可图示如下：



计算机能够机械而准确地按预先规定步骤完成各项运算操作，不会像人那样在运算过程中产生这样或那样错误，因此工作精确程度较高，但这并不是说执行程序所得结果也必定是精确的。事实上，计算机解题时存在有近似性，这种近似性是由量化误差及近似的算法引起的误差所造成的。□

7 计算机仿真及其应用*

一 仿真的意义及目的

建立系统的模型，然后在模型上进行试验，这一过程称为仿真。这里所谓系统包括所有工程和非工程的系统。

仿真的目的：在系统研制之前用于规划、评价和研究；通过系统仿真可以估价系统某一部分的性能，可以估价系统各个部分或各个分系统之间的相互影响，以及它们对整体性能的影响，可以比较各种设计方案，从中获得最佳设计；在系统研制中间用于设计和精密分析，可以对一些新建的理论、假设进行检验；在系统研制成功之后则用于考核设计和训练系统的操作人员等。

早期的仿真主要是物理仿真（或称实物仿真）。物理仿真的优点是直观、形象化。但是要为系统尤其是十分复杂的系统构造一套物理模型，投资很大，周期也很长。另外，在物理模型上做试验，很难修改系统的参数，改变系统结构比较困难。至于社会、经济现象和生态系统，就更无法用实物来做仿真试验了。故现在广泛采用的是数字仿真，即把研究的对象（系统）的主要特征或输入、输出关系，抽象出一种数学表达式，即所谓数学模型。

在某些系统的研究中，还把数学模型与物理模型以及实物联合在一起进行试验，这种仿真称为数学—物理仿

* 本文原载于《自动化博览》1996 年第 3 期。

真，或称为半实物仿真。

由上可知，计算机仿真包括三个要素：系统、系统模型与计算机。联系这三要素的有三个基本活动：模型的建立、仿真模型建立及仿真试验（运行）。

计算机仿真主要研究数字仿真方法、仿真语言、仿真技术、仿真计算机及其应用。仿真方法是研究仿真算法、仿真模型的建立、仿真模型的误差及仿真算法的选择等；仿真语言是研究仿真的程序设计，它们是在高级语言的基础上建立起来的，近年来已有几十种仿真语言问世；仿真技术是研究并行处理的全数字仿真技术和模拟仿真中的寻优技术；仿真计算机则是研究仿真专用计算机的结构与特点。

二　仿真的主要类型

仿真分类方法很多。

1. 根据被研究系统的特征可以分为两大类：连续系统仿真及离散事件系统仿真。连续系统是指对那些系统状态变量随时间连续变化，其基本特点是能用一组方程式来描述。离散事件系统则是指那些系统状态只在一些时间点上由于某种随机事件的驱动而发生变化的系统。在两个事件之间状态变量保持不变，也即是离散变化的，所以称为离散事件系统。这类系统的数学模型一般很难用数学方程来描述，通常是用流程图或网络图来描述。

2. 按使用的计算机分类，则有：① 模拟计算机仿真。由于模拟计算机能快速地解算常微分方程，所以当采用模拟计算机仿真时，应设法建立描述系统特性的连续时间模型。由于在模拟计算机上进行的计算是并行的，因此运算速度快。当参数变化时，容易掌握解的变化，这些是主要优点。主要缺点是当处理多变量时或非线性较强的场合，对于偏微分方程难以求得高精度的解。② 数字计算机仿真。20 世纪 60 年代后，由于数字计算机的发展，它已逐渐取代早期采用的模拟计算机，而成为仿真技术的主要工具。它适用于把数学模型当作数字计算问题，用求解的方法进行处理，而且由于数值分析及软件的发展，使数字式仿真领域不断扩大。由于数字计算机不仅能解算常微分方程，而且还有较强的逻辑判断能力，所以数字式仿真可以应用于更多领域。如系统动力学问题，系统中的排队、管理决策问题。其主要缺点是计算速度不如模拟式仿真。但近年来已开发了大量数字仿真的软件，因而提高了仿真工作的自动化程度。③ 混合计算机仿真。这是一种将模拟式仿真与数字式仿真的优点结合起来，通过一套混合接口（如 A/D，D/A 转换器）组合在一起的混合计算机系统。它兼有模拟

计算机的快速性及数字计算机的灵活性。它不仅能解决系统的动力学问题，而且也能解决许多排队、管理决策等问题，并且还能求解包括流程图形式的模型。这种仿真的结果是模拟模型和数字模型的最优系统。混合式仿真最近也应用于解偏微分方程和求最优解的问题，其缺点是造价昂贵，难以在民用部门推广。

三　计算机仿真的步骤和仿真技术

计算机仿真，概括地说包括“建模—试验—分析”这三个基本部分，即仿真不仅是单纯的对模型的试验，而且包括从建模到试验到分析的全过程。因此进行一次完整的计算机仿真应经过以下步骤：① 明确仿真对象(系统)。要明确以什么样的精密度来仿真对象的哪一部分和仿真什么样的行为，并根据仿真的目的确定所研究系统的边界及约束条件，以及系统的规模及变量个数等。② 建立数学模型(或流程图)。建立什么样的数学模型与建模的目的有密切的关系。如果仅仅要求了解系统的外部行为，则要设法建立一个描述系统的外部行为的外部模型；如果不仅要了解系统的外部行为，还要求了解系统内部的活动规律，就要设法建立一个描述系统输入集合、状态集合及输出集合之间关系的模型，称为内部模型或状态模型。③ 模型变换。即把数学模型变成计算机可以接受的形式，称为仿真模型。④ 设计仿真试验。譬如利用数学公式、逻辑公式或算法等来表示实际系统的内部状态和输入输出的关系。⑤ 模型装载。把模型装入计算机。⑥ 仿真试验。模型装入计算机后，便可以利用计算机对模型进行各种规定的试验，并测定其输出。⑦ 试验结果的评价和分析。首先要确定评价标准，然后反复进行仿真，对诸次仿真的数据进行分析、整理，从代替方案中选出最优系统或找出系统运用的最优值，列出仿真报告并输出。

从仿真技术的观点来看，以上七个步骤可分成三个主要阶段：① 建模阶段；② 模型变换阶段；③ 模型试验阶段。

建模技术、仿真算法、仿真软件技术合起来，就组成一个完整的仿真技术。

四　现代仿真计算机——微型计算机仿真时代

20 世纪 50 年代仿真的主要工具是模拟计算机，60 年代是混合计算机的黄金时代，70 年代后期利用超大规模集成电路研制成的微型计算机，体积小，功耗小，价格便宜，功能也较齐全，因而在仿真技术上得到了广泛的应用。现在，微

型计算机已成为系统仿真的主要工具。特别是 80 年代以后，越来越多的训练仿真系统，采用微型计算机作为仿真计算机。

五　仿真技术的应用

仿真技术是世界发达国家十分重视的一门高新技术，广泛应用于航空、导弹、原子能、宇航等控制系统。由于这些系统比较复杂，成本高昂，加上安全性的考虑，所以采用仿真技术后收到明显的效益。随着仿真的发展，在化工、冶金、电力、汽车制造等工程系统中，仿真技术也得到了广泛的应用。训练操作人员也是仿真技术应用的一个重要方面，比如，训练飞行员的飞机仿真器，训练宇航员的宇航仿真器，训练潜艇驾驶员的潜艇仿真器等。

随着计算机科学的发展，计算机仿真的应用会更加广泛。□

8 计算机网络及其应用*

随着计算机技术的飞速发展，计算机应用的范围亦日益广泛、深入。为完成大量的数据或信息处理、传递、存储和计算任务，就需要把多台计算机用通信线路连接起来，形成计算机网络。关于计算机网络，到目前为止仍没有一个公认的统一定义。有人从应用和资源共享的目的出发，把计算机网络定义为将地理上分散的，以能共享资源（计算机硬件、软件和数据库等）的方式连接起来，并各自具备独立功能的计算机系统的集合。有人从通信的角度来定义计算机网络，认为它是用通信线路和通信设备，将分散在不同地点并具有独立功能的多台计算机系统互相连接，按照网络协议进行数据通信，实现共享资源的计算机集合。这个定义表明，计算机网络是在协议约束下，通过通信系统来实现计算机之间的连接。还有一种定义是：计算机网络是计算机科学技术和通信技术相结合的综合信息系统，也即把远距离及时传输的通信科学技术与科学计算、数据处理、存储技术等计算机技术相结合，从而及时提供各种经过处理后的信息给人或机器。这个定义在一定程度上概括了前面两种定义，比较全面，所以被广为采用。

根据计算机网络的以上定义，计算机网络应具有以下功能：① 数据传递。在计算机网络内最基本的功能之一是终端与计算机、计算机与计算机之间能够相互传递数据和交换信息，能够对地理上分散的单位和业务部门进行实时集中控制与管理。② 共享数据，共享数据库。计算机网络

---

* 本文原载于《自动化博览》1995 年第 3 期。

使大量的、分散的数据能够正确地集中、分析和处理，或使网内各计算机之间以不同方式交换数据，以充分利用网内数据资源，为用户提供方便。③ 资源共享。建立计算机网络的主要目的之一是实现“资源共享”。所谓资源，主要包括软件及硬件资源。网内共享的软件资源包括各种语言处理程序、服务程序和应用程序等。④ 提高可靠性。计算机网络内各站点的计算机连接成网后可以互为后备，当某站点的计算机发生故障时，可以由其他站点的计算机来处理。⑤ 分担负荷。当网内某站点的计算机负担过重时，可将新的作业任务传递到网中任务不饱满的计算机去处理，从而减少用户的等待时间。此外，对于综合性的大型问题，可以采用合适的算法，将任务分散到不同的计算机上进行分布处理。

目前世界上许多国家，除在军事上使用计算机网络系统以外，在民用上如银行、铁道、商业管理、气象预报、飞机订票、医疗等方面，计算机网络也得到了广泛的应用。如国际民航订机票系统，可以自动处理计算票价、开票、货物管理、代订旅馆、膳食服务等。在所有纳入这个网络的城市中，均能在数分钟内订到去任何国家的通航城市的联运机票。这个系统将自动在各公司之间按运送里程转账。例如，一位德国柏林的旅客，在他下塌的旅馆里，通过终端就可以订下一周从英国伦敦飞往美国纽约的飞机票，只需 1～2 分钟的时间，就可以知道他所预订的机票的班次及座位号码。用计算机网络来管理航空系统的售票工作，不仅大大方便了世界各地的旅客，而且更重要的是提高了航空系统的办事效率和运输能力。再如美国贝尔(Bell)研究所的计算机网络，存储着 25 个图书馆的图书、资料，可随时为分布在 8 个州的数万名读者提供资料。当读者通过终端提出某一方面的资料时，只要提出主题词(如书名、作者等)，计算机就可以根据读者要求，在几分钟内就显示或者印出文献资料的内容，既迅速又准确。在很行业务方面，欧美一些国家及日本，早在 20 世纪 60 年代就用自动出纳终端取代了储蓄所出纳员的业务，80 年代实现了全部银行业务自动化和工程化。这些国家各银行之间每天经通信网络流动的货币额高达数千亿美元，每项业务平均办理的时间不到一分钟，而且保证做到准确、可靠、安全、保密，大大缩短了资金周转期，提高了劳动生产率。现在大部分企业和机关已改变了过去的工资发放办法，人们不再把大量现款装到口袋里，而是凭信用卡随时支取，由开户银行转账，这就从根本上改变了货币流通方式。目前，美国一些大型的商用计算机网络已扩展到加拿大、日本和西欧一些国家，国际气象台还建立了“国际气象情报系统计算机网络”。

计算机网络经历了一个从简单到复杂、从初级到高级，不断完善的发展历程。为了在计算机上增加通信控制功能，使远地结点的终端设备通过通信线路直接与计算机相连，形成计算机终端系统，这种系统中使用的通信手段可以是

点一点、多点、公用通信线路以及专用集中器等设备。由于这种系统中仅有一台主机，故它还不是真正的计算机网络。

随着计算机系统的发展，某计算机系统的用户希望使用其他计算系统的资源为其服务，或希望与其他计算机系统联合起来共同完成统一的任务，这就形成了以共享资源为主要目的的计算机网络。这种计算机网络的显著特点是在全范围内必须有共同遵守的网络协议。为此，网络中每个结点的计算机必须有适应网络的支持软件。而且在计算机网络中，计算机之间的通信的管理，通常并不由单一的操作系统实现，即网络中各结点计算机是平等的、独立的，没有主从关系。同时，网络中计算机之间的通信一般是通过通信线路连接的，它们之间的通信交换一般采用串行传输。

随着计算机网络的形成与发展，出现了从逻辑功能上把数据处理与数据通信分开的趋势。这种网络由通信子网和资源子网两级组成。例如，美国国防部高级研究计划局建立的 ARPA 网，就是第一个较大规模的两级计算机网络。它利用租用的通信线路，把分布在各地的通信处理机连接起来构成通信子网，专门负责全网的通信工作，然后把各种资源与通信子网相连，构成资源子网，从事数据处理业务。

目前，计算机网络可分成两类：一类是单处理中心网络，即由一个计算机系统与多台终端设备连接而成的，称为面向终端网络，以传送信息为主要目的；另一类为多处理中心网络，是计算机系统之间互相连接而成的，称为计算机—计算机网，它的特点是以资源共享为主要目的，使应用形式多样化，进一步显示了计算机网络的各种优点。

计算机网络的出现，对计算机的环境变化、设计、操作系统都产生深刻的影响，从而也促进其标准化的早日到来。□

9 办公自动化*

办公自动化(office automation，简称 OA)，是在计算机技术、数据通信技术日益成熟的基础上，为满足社会信息化的需要，正在发展中的一项现代管理技术。它的目的是要提高办公人员的工作效率和工作质量，推动管理工作跨进科学管理和信息决策水平的新时代。广义的办公自动化技术，不仅包括事务处理的自动化，而且也包括现代化信息管理。本文系指后者。

现代化的信息管理，提高了信息的利用价值，而办公自动化就是现代化信息管理的具体体现。据 20 世纪 70 年代中期的统计，世界上已有的计算机，约有 80%用于企业管理和数据处理。进入 80 年代以来，随着价格低廉、体积小、重量轻、功耗小、功能齐全且可靠性较高的微机系统的迅速发展，各种个人计算机(PC)产品如雨后春笋，为实现办公自动化从物质上提供了新的有力保证。

办公自动化是继自动控制、数据处理之后的一个新的发展动向，它是计算机网络技术实际应用的结果，日益受到各国政府的高度重视。目前，办公自动化的应用已日趋社会化。如法国巴黎戴高乐机场是一个现代化的大型国际机场，现已实现了办公自动化：在出入境管理方面，海关人员可将旅客的护照号码、航班和座位输入计算机，管理人员在办公桌上就能直接向中心计算机了解护照的详细内容，诸如护照是否过期、有无嫌疑或被通缉，以及是否重点检查对象等等，减少或避免了排队查护照的麻烦。在大

* 本文原载于《自动化博览》1994 年第 3 期。

商店或超级市场，办公自动化系统能使数十万种商品的进、销、结账变得清清楚楚，大大加速了商品的销售量和资金的周转。在美国，内阁事务办公室是白宫的“神经中枢”，过去这个办公室只有两种基本办公设备，一种是秘书用的电传打字机，另一种是电话机，对付不了神经中枢所担负的事务。1981 年，他们开始引入办公自动化设备，即用先进的多功能设备更新了原来的办公设备，使这个神经中枢取得了惊人的进步。例如，他们利用文字处理设备能快速进行文字处理，及时为总统提供从内阁会议到玫瑰花园举行仪式时所需要的一切文字材料。把这些设备连接起来组成一个网络，在办公室内部、各部门和各机构间进行通信，通过电子邮件传递各部门和各行政机构之间的信息，包括通知议事日程和议事项目、安排会议、变动预定计划、分发书面简报等。特别是在发生紧急事件的情况下，总统发表的声明能在几分钟内通过电子邮件同时分发到内阁成员、各部门和各机构的负责人手里，以便他们发表演说或出席会议谈及当天紧急事件时有所依据。美国 1981 年就有约 70％的办公室用上了文字处理机及传真设备。据 IBM 公司估计，1980 年大约 50 名办公人员使用一个终端，到 1990 年大约 2 名工作人员就使用一个智能终端。由于数据终端与工作人员的比值是用来衡量一个社会信息化程度的重要标志，因此这个数字说明美国办公室自动化有相当高的程度。另外，美国办公室个人计算机以惊人的速度增长着，它为美国办公室自动化提供了雄厚的物质基础。日本继美国之后，也从美国引进办公自动化设备，极大地提高了办公效率、劳动生产率和企业竞争力。

目前的办公自动化主要是指在一个大的部门或企业中，通过计算机网络，把办公机器与工作人员结合起来构成一个有机整体，人们能够共享系统的资源，可以根据自己的目的和要求及时地进行信息加工和信息处理。这里，进行信息处理和利用信息资源的当事人，已不再像过去那样仅限于计算机专业人员或专家，而是适用于办公室的所有工作人员。

办公自动化设备一般包括：事务处理计算机、信函机、信息传视机、传真机、电传打字机、电子文件处理机、智能电话等。

办公自动化系统是一个计算机网，目前多是局域网。电子邮件是办公自动化系统传输信息的重要手段和途径，使用者把发送的文件，通过终端送入系统，然后按指定要求传送给接收人，从而实现信息的即时传送。由于传输渠道畅通，减少了中间层次，信息得以及时上通下达，从而明显地提高了工作效率，使经营管理的工作方式得到很大改善。

80 年代后，办公自动化发展迅速，逐渐成为一门综合性很强的新兴学科。虽然它涉及的关键技术很多，但都是以管理好庞大的信息系统为出发点，都离不开计算机。一个较完整的办公自动化系统，必须包括信息收集、信息处理、信

息传递和信息存储等基本环节。为采用现代化手段解决这些基本环节，必须解决以下一些关键技术问题。

1. 数据处理。数据处理也就是我们通常所说的数据分析。对大量的信息（数据）进行搜集、转换、分类、组织、计值、存储、检索、维护等，有时根据需要还要绘制数据分布曲线或印出报表，或根据大量的数据进行预测等。这些处理工作一般不涉及复杂的数学问题，但数据量大，时效性强，故必须用计算机完成。

数据库是存放数据的仓库，数据库管理系统实现了在数据库环境下数据的存取、检索等功能。数据库管理系统广泛应用于国家自然资源管理、情报检索、航空业务、企业管理、银行业务等。例如采用计算机进行情报检索，一分钟可检索 1 800 篇文献，两个小时就能查阅五个专业的全部资料。当读者通过终端提出需要某一方面的资料时，只要给出主题词（如书名、作者等），在几分钟内计算机就可搜索出有关方面的几十、几百篇文献目录，并且计算机还可根据读者要求，显示或打印文献资料内容。

2. 网络通信技术。办公自动化要成为现实，关键的技术是把各种信息处理设备连成网，使它们能互相通信和共享资源。由于以计算机为基础的各种信息处理设备日益增多，数据通信正在迅速增加，需要发展快速的局域网络。局域网的地理范围是小规模的，最普通的是适用于单一办公楼内。局域网的扩展性好，灵活方便，有可能连接多个计算机厂家生产的设备。它支持微型计算机、小型计算机、大型计算机、终端机、外围设备和其他数据处理设备。在办公自动化系统中，这种网络不仅可以传送数据，而且可以传送声音、图像等信息。在设备种类繁多和涉及的通信方式混杂时，局域网可能是较为理想的选择。

局域网对内可实现办公事务处理自动化，提高办公效率；对外可经由调制解调器与电话线路相连，通往当地计算中心，连接到公共数据网，进一步与社会各界交换信息，共享网上的硬件、软件及数据资源。

近年来，一些国家都在努力发展电缆电视、会议电视、可视电话、可视数据和传真等，这对办公自动化起着积极的推动作用。

3. 人机分工。办公自动化系统是一个人机系统，在这个系统中，某些信息处理工作要求计算机完成，而另一些工作则由管理人员或操作员来完成。由于计算机技术的飞速发展，致使越来越多的人工职能可由计算机在更大的范围内、更迅速、更准确地完成。但是，人的智慧、知识和经验，比起计算机来说具有某些固有的优越性。人可以从不完整的信息中推断出完整的信息，人还可以在一种试探的环境中进行自己的工作，并且人能够用自己的全部经验解决所遇到的问题。与此同时，人还需要使自己的工作具有进取性和多样性，而使自己不断进步。这就意味着使用一种技术上最理想的分工方案，对操作员和经理来说

所构成的工作环境可能不一定是最合适的。因此，人机分工就是要把最适合于人做的工作和最适合于机器做的工作区别开来，以做到人尽其才，物尽其用，并使人和机器协调工作，充分发挥整体优势，提高办事效率。

在国外，办公自动化的一个新的发展方向，是打破传统的许多人挤在一起办公的方式，实行居家办公。

在我国，国务院于 1985 年成立“办公自动化专业领导小组”，并于同年召开了全国第一次办公自动化专业会议。会议确定了我国实现办公自动化的指导思想，即办公自动化必须为国家经济建设服务。办公质量的高低，直接影响生产力的发展，因此，在我国实现办公自动化，对促进机构改革、缩小行政部门编制、提高办公效率，具有重大的现实意义。办公自动化是对传统管理方式的挑战，是经营管理的一次革命。□

10 即将崛起的光计算机

自从1946年第一台电子数字积分计算机ENIAC问世以来，从元器件方面说，计算机的发展已经历了电子管、晶体管、集成电路、大规模和超大规模集成电路共计四代的变革。但从整机原理和设计思想来说，都是按美国数学家冯·诺伊曼提出的ENIAC的方案设计并制造出来的。因此，这种类型的计算机又称为冯·诺伊曼型计算机。

尽管冯·诺伊曼型计算机还是计算机发展的基础和主流，然而，由于冯·诺伊曼型结构是固有的串行结构，在此基础上发展起来的高级语言反映了这一模型逐字加工的特点，因此开发其中的并行性有其致命的弱点。同时，在冯·诺伊曼型基础发展起来的软件系统已越来越复杂、庞大，其正确性无法保证，加之软件生产率低下等问题，构成了所谓的"软件危机"，使得计算机系统的整个性能大幅度下降。因此，计算机体系的设计者不断地寻求各种途径更新计算机结构，从根本上摆脱这种传统式的计算机模型，尽可能地开发并行结构，提高计算机的处理性能。于是，一种全新概念的非冯·诺伊曼型计算机便应运而生，并逐渐发展起来，这就是所谓第五代计算机。

光，在整个物质世界中占有特殊的地位。所谓光，通常是指照耀在物体上，使人能够看见物体的那种物质。例如太阳光、月光、灯光等。可见光是波长为7.7×10^{-3}厘米的电磁波。由于光是电磁波的一种，所以光也被称为光波。随着新技术革命的发展，现代光学已渗入到光传感器、激光、光纤通讯、光开关、光存储等先进技术领域，成为新技术革命的一支强劲生力军。科学家预测，随着现代光

学与计算机技术、微电子技术相结合，在不久的将来，光计算机不仅将会出现，而且大有替代电子计算机之趋势。

光计算机又称为第五代计算机，是利用光的一些基本性质而研制的一种计算机，因此它具有电子计算机无法比拟的优点。如：① 具有并行传输和处理信息的能力。同一光波导(用以构成集成光路中的连接元件)能并行地传输很多不同波长或不同偏振态的光波，它们之间不会发生干扰。因为两束光要发生干涉，必须是波长相同、振动方向一致(同一偏振态)和有确定不变的初始位相差。② 可采用非二进制，提高运算速度。光计算机与电子计算机一样，可采用双稳态器件，实现二进制；而光器件已突破双稳态，出现三稳态器件。最近日本研制成功的光多稳态元件，则可采用三进制、四进制。同一个数，其位数往往比二进制少。如一个数 3(在十进制下)，二进制为 011，三进制为 010，四进制为 003。位数少可提高运算速度。③ 无时钟歪斜、摆脱冯·诺伊曼"瓶颈问题"、寄生电容限制响应速度和互联带宽限制等问题。如模拟光计算机原理，它的主要特点是能高速并行地处理数据，它克服了电子计算机的冯·诺伊曼瓶颈问题、寄生电容限制等四大固有缺点。由于目前的电子计算机是根据冯·诺伊曼提出的基本原理设计的，其中最重要的是采用二进制和串行结构，以尽量减少互连线的数目，但延长了传输时间。它好似一个大容量的瓶子，但瓶颈很细，进出速度较慢，即所谓"瓶颈效应"。电子计算机内有大量的电子元件和连线，每个元件均有一个响应时间。例如最重要的双稳态(0—1)触发器，由于电阻、电容和寄生电容存在，不可能实现 0→1(或 1→0)的瞬时转换，响应有一个滞后效应，许多平行连线间存在分布电容，因此传输亦有一个 RC 延时常数。另两个固有缺点是时钟歪斜和互连带宽限制。当然，模拟光计算机也存在一些缺点，主要是精度不高(只有 8～10 比特)、动态范围不大、通用性差，这些缺点限制了它的应用领域，于是人们自然提出了数字光计算机。它既有模拟光计算机的高速并行处理的特点，又有电子计算机的精度高、通用性强的优点。近年来，数字光计算机用的逻辑器件、存储器件、通信和互联、算法和结构等方面都有着很快的发展。不久的将来，新一代的超级数字光计算机一定能够走向市场，满足信息社会的需要。

光学计算是指利用光束进行数据计算和资料处理。因为它具有时间上的高速性、空间上的并行性以及不受电磁干扰等优点，因此在国际上掀起了一场光计算机热潮。光速极快，光速开关的转换比电子开关快几千至几百万倍。光信号之间可毫不干扰地沿着各自通道或并行的通道传送，然后用全息的或图像的方式存储信息，从而使光计算机能并行处理数据并大大增加容量。要使光计算机变成现实，研制出实用的光路器件就成了问题的关键。

1. 光开关器件。众所周知，晶体管对电信号具有放大和开关作用，是一种广泛应用的电子器件。电子学家仿效晶体管的功能研究光放大、光开关的机理，经过十多年的努力，光晶体管已在实验室初步研制成功。这种光晶体管是一种不用把光信息转变成电信号而直接对光信息进行处理的光路器件。它是利用折射率随入射光束强度变化的非线性晶体的性质，使入射光束的强度产生极大的改变，如同晶体管从基极流向集电极电流的微小变化引起从发射极流向集电极的巨大变化那样，使之具有光放大作用，也可类似晶体管开关电路那样构成晶体管开关光路。英国罗特—瓦特大学教授史密斯和日本京都大学教授佐佐木昭夫分别于 1982 年和 1983 年各自独立地研制出光晶体管。

2. 光存储器件。光存储器件是一种以光的形式存储信息，由化合物半导体材料组成的光路器件。光电子学专家通过对入射光强与透射光强之间相互关系的深入研究，发现了光的双稳态性质，从而研制出了一种光双稳态器件。这是一种处于激光脉冲工作状态的存储器件，它的存储功能是根据激光脉冲是否输入，由固定稳态及稳态的转换来区别所存的“1”和“0”的信息的。日本对光双稳态器件做了鉴定试验，结果表明存储器错误率小于 10^{-10}，也就是说在 10^{10} 个信号中产生错误的还不足一个，充分说明了这种光存储器件的有效性。

我们知道，电子计算机是在电子器件及其构成的线路中利用电信号进行运算、传输、存储和信息处理的，而光计算机则是一种由光信号进行数字运算、逻辑操作、信息存储和处理的新型计算机。与电子计算机使用的集成电路相似，光计算机的基本组成部件是集成光路。从 20 世纪 70 年代起，在光学领域兴起了一个新的分支学科，叫做集成光学，这是一门专门研究、设计和制造集成光路的科学。运用集成光路技术，可以把光开关或光存储等器件集积在一块芯片上，制成单一功能的集成光路，也可以将光源、光波导、光开关、光存储等器件集积在一块芯片上，以组成一个完整的光系统。如果要得到一台光计算机，就可以选用集成光路进行组装，而光导纤维则能用作光计算机之间的直接通讯线。

众所周知，晶体管可以用作高速开关，但由于惯性作用的影响，其开关速度并不是可以无止境提高的。目前使用的开关速度最快的硅晶体管，改变状态的开关时间不可能低于 0.4 毫微秒，而光晶体管理论上将达到 1 微微秒的水平。这种光开关器件每秒钟能进行 10 000 亿次逻辑运算，与当前运算速度高达 25 亿次的巨型计算机相比较要快 400 倍。光计算机具有较强的并行处理的能力，它可以同时处理几路信息，其原因在于几束光通过光晶体管时能够相互保持独立，而几路电流输入晶体管时却会混在一起。另外，电子计算机是利用电子传输信息的，容易受磁场的影响，而光计算机则不然。

目前研制的光计算机主要有三种类型：① 相干光学模拟计算机。主要是由

相干光源、光束扩展器、图像输入装置、傅里叶变换镜头、滤波器、输出装置等组成，它具有处理速度快、信息容量大、可并行处理、对二维图像可直接进行运算等优点。相干光学计算机已经十分广泛地应用于图像改善与增强、图样识别、合成孔径等方面，但是由于整个运算过程是模拟的，因此精度不太高。② 光学一数字混合计算机。这是由相干光学模拟计算机和数字电子计算机组合起来的一种光计算机，它兼有前二者的优点。这种计算机处理图像接近人眼和大脑观察与识别外界事物的过程，是极其有发展前途的信息处理设备。通常用相干光学模拟计算机对图像作预处理，而后用数字电子计算机作精处理。它可用于图像改善与增强、图样识别、编码和成图技术等。③ 全数字化光学计算机。光学模拟计算机的发展将导致全数字化光学计算机的出现。这种计算机可对二维图像同时进行数字化处理。由于它的运算和处理的并行性，因此它的速度和容量至少是电子计算机的平方倍。全数字化光学计算机的发展取决于纤维光学、显示技术、薄膜光学和半导体技术的发展。目前，处理二维信息的数字化的"字"计算机已经出现，尽管它仍需作进一步的改进与完善，但已经显示出无比的优越性。这无疑是计算机技术的重大突破。

1990 年 1 月 29 日，美国贝尔实验室宣布研制出世界上第一台光计算机，它采用砷化镓光学开关，运算速度达每秒 10 亿次。尽管这台光计算机与理论上的光计算机还有一定距离，但已显示出强大的生命力。□

运筹学：用数学方法寻求决策最优化

1 最优化方法及其发展趋势

最优化方法是近30多年来随着计算机的应用而发展起来的应用数学的一个重要分支。所谓"最优"，就是做一件事情要取得最好的效果。今天，人们不论从事什么活动，都希望所采取的策略能使某个或某些指标达到最优。如工业生产中，在不增加设备和费用的条件下，如何选择合适的工艺条件才能达到优质高产？或在确保质量的前提下，如何才能使成本最低？最优化问题在现实世界中广泛而大量存在着。最优化方法就是将最优化问题表示成数学问题，并迅速找出它的最优解的数学方法。在数学上有一类求函数的极小值或极大值的问题，最优化问题往往就是求目标函数的极小值或极大值。用数学语言说，最优化方法就是在一组约束条件下，求目标函数极小值或极大值的方法。因此最优化方法是一种数学方法而不是工程方法，不仅在运筹学、系统工程、经济管理等领域内有着重要的意义，而且在工程设计上也得到日益广泛的应用。

最优化方法得到迅速的发展，这一方面是由于近代工农业生产、科学实验和国防科学的发展提出了大量的优化问题，另一方面也是与计算机的发展和广泛应用分不开的。第二次世界大战后，在工业生产、宇航、国防科学和大系统中出现的上万个甚至上百万个变量的最优化问题并不少见，手算已无能为力，离开了计算机就寸步难行。因此最优化方法既是一种数学方法又是一种计算机方法，两者结合在一起形成了一门新技术。如果给汽车装上微机，使它在各种道路上能最优运行，那么一辆车所节约汽油的价值很快就能抵消计算机的费用。国外许多大城市的公

共汽车和民航的优化调度，不仅都取得了惊人的经济效益，而且提高了服务效率。在今后若干年内，最优化方法将随着计算机网络的发展而进入政治、经济、军事和社会各个方面，微机处理将把最优化方法带进各个角落。

在20世纪50年代，求解最优化问题的数学方法主要是古典的微分法和变分法。50年代以后，对于实践中出现的许多最优化问题，古典的方法已不能奏效，从而产生了线性规划、非线性规划、贝尔曼的最优化原理和动态规划、庞特里亚金的极大值原理以及卡尔曼的最优滤波器，这些构成了现代最优化方法的基础。但这些古典的微分法、变分法等主要是解决目标函数能用解析式表达的最优化问题。有时目标函数十分复杂，上述各法都无能为力，而且实践中有大量这样的最优化问题，甚至找不出具体的函数表达式。它的导数很难求出，或根本不存在，而它的数值只有通过实验才能确定，这样的最优化问题称为“实验最优化问题”。解决这类问题的方法要求以较少的实验次数找出问题的最优解。如0.618法（或称黄金分割法），是一种常见的单因素优选法。70年代，华罗庚教授带领小分队在全国大中城市的工矿企业推广优选法，取得了巨大的经济效益，主要就是推广应用了0.618法。正交优化法（也称正交法），是一种多因素优选法，这种方法广泛应用于技术革新、工艺改革、产品设计和科学实验等方面。这些领域的每个项目一般含有多个条件（称为因素），因而要做多因素实验，以找到最优方案。如何处理多因素问题以尽快取得最优生产方案，历来是一大难题。正交优化法利用一种现成的规格化的正交表，只需做较少次数的实验，就能找到最优方案。例如沈阳风陵机械厂高温点焊胶配方实验组在一年内做了近900次实验，各项指标均未过关。后来改用正交优化法，只用了一周时间做了16次实验，就找到了最优化配方，各项指标都达到国家标准。据1987年有关方面统计表明，我国应用正交优化法进行了数万项实验，获得经济效益近10亿元，其中仅北京市化学工业总公司用正交优化法做实验100多项，取得经济效益达2 000万元。我国数学家从理论上对正交法的高效率作出了解释，他们的研究成果受到国际同行们的高度重视。

当最优化问题的目标函数或约束条件中至少有一个是自变量的非线性函数时，则称这种最优化问题为非线性最优化或非线性规划问题。解非线性规划问题要用非线性规划的方法，它比解线性规划问题更难，因为解非线性规划问题不像解线性规划问题有单纯形算法这一通用方法。到目前为止，还没有建立起适用于各种非线性规划问题的一个通用算法。求解非线性规划问题的各个方法都有自己特定的适用范围，因而这是需要人们深入研究的一个领域，但由于很多问题需要进一步精确化，以及计算机的发展，使非线性规划在近二三十年得到迅速的发展，开始形成了最优化问题的一个重要分支。它有广泛的应

用，特别是为最优化设计提供了数学的理论基础。

我们知道，如果线性规划的最优解存在，则最优解只能在可行域的边界上达到，特别是在可行域的顶点上达到；而非线性规划问题的最优解如果存在，则可能在约束集合形成的边界上，也可能在可行域以内的任一点达到。由于线性规划的目标函数为线性函数，因而求出的最优解就是整个可行域的全局最优解。非线性规划则不然，有时求出的某个解虽是一部分可行域上的极值点，但并不一定是整个可行域上的全局最优解。只有在一定条件下，局部极值才是全局的极值。

求解非线性规划的方法目前大体上分成两类：① 解析法。当目标函数有明显的解析表达式时，一般用求导数的方法（或变分法），求出最优化的必要条件，再通过必要条件将问题简化。由于用了函数的解析性质，故称为解析法或间接法。拉格朗日的乘数法、古典微分法、变分法、庞特里亚金的极大值原理、卡尔曼的最优滤波器、动态规划等都属于解析求优法。1951 年，英国数学家库恩和塔克提出了处理不等式约束条件下的非线性最优化问题的完整理论，使非线性规划理论发展到一个新的阶段。② 直接法。当目标函数较复杂或不能用自变量的显函数表示时，就难以求最优化的必要条件，这时要用直接搜索方法，经过若干次迭代搜索到最优点。这种方法常常是根据经验或实验得到的，故称为直接方法或实验最优化方法。直接法又分为两部分：一部分是同时做实验来求优，这部分属于设计的范畴，如正交实验设计和调优运算；另一部分是序贯实验法，此法简单说就是把每次实验结果与上次所得的结果中“最好的结果”进行比较，然后根据某种策略决定下一步的实验。序贯法目前已有不少行之有效的方法，总起来可分：① 区间消去法，又称为一维搜索。这是求单变量函数无约束极值较为有效的方法，也称为线搜索，是沿某一方向寻求目标函数极小点的方法，如 0.618 法、斐波那契法、插值法等。② 爬山法。这是多变量函数的直接寻优方法。这种方法的实质是利用已有信息，逐步改善目标函数，最后达到最优点。其中包括坐标轮换法、步长加速法、共轭方向法、单纯形法（是一种多维直接搜索法）、随机搜索法等。此外，以梯度法为基础，一种解析与数值计算相结合的求优法也属于直接法的范畴。□

2 线性与非线性理论*

形式化、数量化(有时还要最优化)的研究是现代科学技术的突出特点。控制论也不例外。因此研究控制系统的第一步,就是要建立它的数学模型,也就是要把一个控制系统用数学表达式加以描述。数学模型是研究自动控制系统的基本依据,没有一个确切的数学模型,就不可能对系统进行定量的分析研究。

由于控制系统具有动态的特点,各个变量在动态过程中的关系当然总是一般表现为微分方程形式。在微分方程分类中,有线性和非线性微分方程。当一个控制系统的数学模型是线性微分方程时,称为线性系统,它必须满足两个条件:① 齐次性。若输入 $x_1(t)$ 引起输出为 $y_1(t)$,则以常数 a 作用于输入 $x_1(t)$,使新的输入为 $ax_1(t)$,则输出相应地为 $ay_1(t)$。② 迭加性。若分别输入 $x_1(t)$,$x_2(t)$,对应的输出分别为 $y_1(t)$,$y_2(t)$。现将 $x_1(t)$ 与 $x_2(t)$ 的迭加式 $x_1(t)+x_2(t)$ 作为输入,则输出为相应的迭加式 $y_1(t)+y_2(t)$。

反之,当一个系统不满足线性条件中的任何一条,或者说,在一个自动控制装置中只要包含了一个非线性的环节(元件或部件)时,就是非线性系统。一般说来,实际的自动控制系统都是非线性的,只是当我们建立其数学模型时,往往把一些次要的非线性因素加以忽略,使输入与输出之间的关系成为线性的数学模型,即所谓小偏差线性化。这时用线性微分方程代替非线性微分方程,用线性理

---

* 本文原载于《自动化博览》1997 年第 5 期。

论来进行分析，所得的结果在实践中证明大体是正确的。我们知道，对于非线性方程求解是比较困难的，除了对某些情况可以找出一些办法使之线性化以外，大多数只能用近似的数值计算来处理，甚至只能作定性的描述与讨论。

近年来，线性系统理论逐渐发展成为现代控制理论的一个独立分支，而且是非常重要的一个分支，也是自动控制理论基础最成熟最实用的部分。其中包括一些奠基性的基本概念，例如能控性与能观测性的概念，状态反馈和状态估计的概念等；也包括线性控制系统分析与综合的基本方法，例如极点配置问题、解耦问题、实现问题及状态观测器的设计问题等等。此外，由于电子计算机应用的迅速发展，使一些古老的理论获得了新生。例如古典控制理论中的频率法被推广应用于线性多变量系统的分析与综合，这被称为“现代频域法”，进而编制了各种典型的计算程序，形成了一套完整的应用软件。再如李亚普诺夫（A. M. Lyapunov）稳定性理论，也由于电子计算机的应用而获得了新生，在自动控制系统的分析与综合方面，发挥了新的作用。

在建立了控制系统的数学模型之后，为了对它求解，需要一些方法。特别是当微分方程比较复杂时，如阶数较高，求它的特征方程就较困难，因此就必须找到某些简便的方法来求解微分方程。拉普拉斯变换的方法便是其中著名的一种，简介如下。

定义 设函数 $f(t)$，当 $t\geqslant 0$ 时有定义，而且广义积分 $\int_0^{+\infty} f(t)\mathrm{e}^{-st}\mathrm{d}t$（$s$ 是一个复变量）在 s 的某一域内收敛，则由此积分所确定的函数可写为

$$F(s)=\int_0^{+\infty} f(t)\mathrm{e}^{-st}\mathrm{d}t。$$

称上式为函数 $f(t)$ 的拉普拉斯变换式（简称拉氏变换式），记为 $F(s)=L[f(t)]$，称 $F(s)$ 为 $f(t)$ 的拉氏变换。L 为拉氏变换的符号，起着拉氏变换的作用。

拉氏变换有一些重要的性质，对于控制系统的分析与处理是很方便的。例如，一个线性微分方程，经过拉氏变换后，就把这个微分方程变为与它等价的代数方程。这样，求解就简便得多了。这是运用拉氏变换法的一个最大的优点。

线性系统虽然便于数学处理，但往往不能较精确地描述系统运动的状态，甚至不能解释许多实践中经常遇到的问题，特别是在设计控制系统时，往往有意识地引入一些非线性环节，用以提高系统的工作质量。因此在自动控制系统中有时仅仅运用线性系统理论是不够的，所以深入地认真研究非线性系统的理论和方法就具有实际的重要意义。由于非线性系统的种类纷繁，又没有普遍有效的数学工具可以利用，因而只能计算其近似值和进行定性的描述。

稳定性是控制系统的一种重要特性，是动态系统研究的首要问题。在经典

控制理论中，对于单输入一单输出的线性定常系统，我们可以用奈魁斯特判据、劳斯判据等来判别其稳定性；但当系统是时变的或非线性的时，这些判据就不适用了。线性系统的稳定性取决于系统的结构参数，与系统的初始条件及外界干扰的大小无关；而非线性系统的稳定性不仅与系统的结构与参数有关，而且与初始条件及外界扰动的大小都有关系。在古典理论中没有给出同时适用于线性与非线性系统的一般的稳定性定义。1882 年俄国数学家李亚普诺夫给出了对于任何系统都适用的稳定性的一般定义。稳定性的物理意义是一个系统的响应是否有界，他据此将系统的稳定性定义三种情况：稳定，渐近稳定，不稳定。李亚普诺夫还给出了研究系统稳定性的普遍方法——李亚普诺夫第一法和李亚普诺夫第二法。第一法是把非线性函数用近似级数表示，然后用近似方法求解这个非线性方程；第二法不是通过解方程，而是通过一个叫做李亚普诺夫函数的纯量函数来判别系统的稳定性。由于第二法不用解方程就能直接判别系统的稳定性，所以又叫李亚普诺夫直接法。

对于现代控制系统来说，李亚普诺夫直接法具有明显的优越性，因为求解非线性方程和时变系统的状态方程一般都很困难，所以不用解方程就能确定系统的稳定性就显得更为优越。另外，用李亚普诺夫第二法不仅能判别系统是否稳定，还能分析线性或非线性系统的瞬时响应。因此在现代控制理论中受到重视的是李亚普诺夫直接法。

李亚普诺夫直接法是从能量的观点来判别系统的稳定性的。如果一个系统储存的能量是逐渐衰减的，则这个系统是稳定的；反之，一个系统如果不断从外界吸收能量，系统的能量越来越多，则这个系统是不稳定的。对给定的一个系统，只要能找到一个正的纯量函数 $v(x)$ 表示系统的能量，如果导数 $\frac{\mathrm{d}v(x)}{\mathrm{d}t}$ 是负的，则这个系统是稳定的。李亚普诺夫直接法就是用 $v(x)$ 和 $\frac{\mathrm{d}v(x)}{\mathrm{d}t}$ 的正负号来判别系统的稳定性，这个函数 $v(x)$ 就叫做李亚普诺夫函数。李亚普诺夫直接法是普遍的方法，李亚普诺夫函数不局限于能量函数。实际上很多复杂的系统往往不能直观地找到能量函数，我们可以把 $v(x)$ 看成是假想或虚构的能量函数，只要找到一个正的纯量函数 $v(x)$，根据 $v(x)$ 和其导数 $\frac{\mathrm{d}v(x)}{\mathrm{d}t}$ 就能判别系统的稳定性。

这样一来，判别系统的稳定性问题就可以归结为寻找李亚普诺夫函数 $v(x)$。过去，要想寻找到李亚普诺夫函数是靠试探，要凭个人的经验和技巧，致使李亚普诺夫直接法长期不能被推广应用。计算机的发展，为我们寻找李亚普诺夫函数提供了有力的工具。应用计算机不仅可以找到所需要的李亚普诺夫函数，而且还能找到系统的稳定区域。□

3 线性规划算法的新进展*

所谓线性规划，就是在一组线性方程和线性不等式的约束下，去寻求使线性目标函数达到极小或极大值的一组未知量，即寻求最优解。它的一个数学表达式是：求出 n 维向量 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$，满足约束条件

$$\begin{cases}\boldsymbol{AX}\leqslant\boldsymbol{B}, & (1)\\ \boldsymbol{X}\geqslant\boldsymbol{0}, & (2)\end{cases}$$

使线性目标函数

$$Z=f(\boldsymbol{X})=\boldsymbol{C}^T\boldsymbol{X}$$

达到最小或最大值，即

$$\min(\text{或}\max)f(\boldsymbol{X})=\boldsymbol{C}^T\boldsymbol{X}。$$

式中“T”表示转置，$\boldsymbol{A}$ 是 $m\times n$ 矩阵，$\boldsymbol{B}$ 是 m 维列向量，$\boldsymbol{C}$ 是 n 维行向量。满足约束条件(1)、(2)的 n 维向量称为 LP 的可行解，所有可行解组成的集合称为可行解集，把目标函数取最小值或最大值的可行解称为 LP 的最优解。因此求解 LP 就是从它的可行解中寻求使目标函数取最小或最大值的可行解。从几何上看，可行解集是约束不等式组成的抽象高维空间中的一个凸多面体，而使目标函数达到极值的向量的终点，一般都在这个凸多面体的顶点上，最优解就是使目标函数达到极值的顶点。

1. 单纯形法

1947 年，美国青年数学家、斯坦福大学教授丹茨格(Dantzig)发明了求解 LP 的通用算法——单纯形法。单纯形法是一种迭代求解算法，其基本思想是：沿着凸多面

* 本文原载于《自动化博览》1994 年第 2 期。

体的某个顶点(初始解)出发,根据某种原则(称为列选择原则)向邻近顶点运动,由一个顶点迭代到另一个顶点,使相应的目标函数的值一次比一次好。由于顶点是有限的,而最优解又一定在顶点上,所以经过有限次迭代,最后一定能求得最优解,或判定它无最优解。由于可以有各种形式的列选择原则,因此就有各种不同形式的单纯形法,如原单纯形法、对偶单纯形法、原—对偶单纯形法等。它们的有效性是举世公认的,所以多年来求解 *LP* 问题,一直都采用单纯形法。实践也证明,只要变量个数不超过 1.5 万或 2 万,单纯形法实行起来确实是有效的。因此,单纯形法已被制成标准软件(或软件包),广泛应用工商业、交通运输、军事、政治、社会等决策领域,起到了重要决策参谋作用。丹茨格本人也因发明单纯形法而获得了美国运筹学会的第一次冯·诺伊曼奖,当选为美国科学院院士,并荣获美国国家勋章。

那么单纯形法是不是多项式的时间算法?回答是否定的。因为 1972 年,美国数学家克利和明蒂以《单纯形法究竟好到什么程度》为题发表文章,文中他们构造了一个具有 n 个变量和 $2n$ 个约束不等式的病态例题,用单纯形法来解这个例题,选择适当的初始点,使得必须检验完可行域的所有顶点,才能得到最优解。结果它的计算次数大于 2^n,这清楚地说明单纯形法不属于多项式时间算法。紧接着,1973 年,杰罗和扎德又分别发表文章指出,所有各种形式的单纯形法全是指数时间的。那么为什么单纯形法在实践中的效果那么好?原来在实际应用中像克利和明蒂构造的那种病态例子,几乎从未遇到。1982 年,联邦德国学者博格瓦特和美国学者斯梅尔用概率的方法证明:若不考虑最坏情况,而仅考虑计算时间的平均值,则单纯形法的平均收敛速度是多项式时间的,甚至可以说是线性的。但实践里出现的问题中,上百万个变量和数十亿个可能解的大型线性规划问题并不少见,即使是强大的快速电子计算机也不可能检验每个解是否为最优解。此外,处理大型问题往往需要付出昂贵的寻优费用。所以,从计算复杂性的角度来讲,单纯形法并不是一种好算法。因此,计算机必须用一种特殊程序,即最佳算法去检验尽可能少的解,从中找到最优解,就成为国际科技界竞相研究的热门课题。

2. 椭球算法

由于在经济管理、工程、通信等方面存在着许多大型复杂的线性规划问题,因此需要一种比单纯法更快、更为有效的求解线性规划的算法,以最大限度地提高解题的效率和尽可能多地降低成本。在此背景下,1979 年,前苏联年轻的数学家哈奇杨(Л. Г. Хачиян)发表了一篇题为《线性规划的多项式算法》的著名论文,提出了一个完全不同于单纯形法的解线性规划的椭球法,并证明了这个算法是多项式时间的。椭球法也是一种迭代求解法,但它不是从可行解集的一

个顶点迭代到另一个顶点，而是向“内部深入”，始终保持在一个椭球上，从较大的椭球缩到较小的椭球。哈奇扬把他的算法用于线性规划，证明了线性规划是可以经过多项式运算解决的问题，而且是多项式时间的收敛。哈奇扬的成果石破天惊，轰动了国际数学界。因为他的算法的主项为 n^6L^2（n 是变量个数，L 表示问题的输入数据的位数），它只有在 n 大到天文数字时才是指数函数。然而，在实际应用中，人们发现椭球法并不比单纯形法更好。但哈奇扬的算法的重要性仍旧存在，这就是他提出了一个解线性规划的多项式时间算法，从而在理论上证明了线性规划属于 P 类问题，解决了多年来人们对线性规划是否属于 P 类问题的怀疑。

那么为什么一个多项式时间算法反不如一个非多项式时间算法的单纯形法有效？现在清楚了，这与线性规划的应用背景有关。线性规划的应用课题通常是大规模的，其系数矩阵有稀疏的特点（即矩阵的系数相当大的部分为 0）。对单纯形法来说，一方面人们在实践中已总结出处理稀疏性的好经验，在使用单纯形法的过程中，稀疏得到了较好的保护。但椭球法则不同，它几乎不能保护稀疏性，因此不能享受稀疏性带来的好处。

3. 卡玛卡算法

卡玛卡（Karmarkar）于 1955 年出生于印度，在美国加利福尼亚大学伯克莱分校获博士学位后，任美国贝尔实验室数学部研究员。他在长期的研究工作中，认识到单纯形法的缺点，并受到椭球法的启迪，终于在 1984 年在线性规划的理论方面取得了突破性进展。他发明了一种解线性规划的新的多项式时间算法，他的算法量级是 $O(n^{3.5}L^2)$ 的，比哈奇扬的算法量级 $O(n^6L^2)$ 大大前进了一步。和哈奇扬的算法一样，他的算法也是从内部逼近最优解，但用的是射影几何原理不断变换的办法。其基本思想是：通过射影变换把卡玛卡线性规划的标准问题转化为在球域上极小化另一个线性函数，求出在球域上的最优解，再用逆变换将此解返回到原决策空间里去，从而得到原问题的一个近似解。重复以上过程，得到点 $\{\boldsymbol{X}^{(k)}\}$，此点列在多项式内收敛于问题的最优解。所以，卡玛卡算法也需要作逐次迭代，但由于属于多项式，因此大大减少了迭代次数，加快了求解 LP 的速度。1984 年夏天，卡玛卡用他的算法同单纯形法的标准软件进行计算比较，在变量个数为 5 000时，只用了单纯形法的$\frac{1}{50}$的时间就给出了答案，即比单纯形法快 50 倍。

卡玛卡的解线性规划的多项式时间算法，是研究线性规划的重大突破，引起了美国及国际科技界的强烈关注，美国各主要报刊都以醒目的标题报道了这一消息。它的有效性终于在 1988 年于日本东京举行的第 13 届国际数学规划会议上得到了公认，卡玛卡本人也获得了这一年的富克逊奖。事隔不久，卡玛

卡又宣布，他的算法比单纯形法不是快 50 倍，而是 250 倍。卡玛卡算法为求解大型复杂线性规划问题提供了新的有力工具。

使用卡玛卡算法，首先要把一般线性规划问题化为解卡码卡算法的标准问题。称如下形式的线性规划问题为卡玛卡标准型

$$\begin{cases}\min \boldsymbol{C}^T\boldsymbol{X},\\ \boldsymbol{X}\in\Omega\cap S。\end{cases}$$

其中 $\boldsymbol{C}^T=(c_1,c_2,\cdots,c_n)$，$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$，$\Omega=\{\boldsymbol{X}\mid\boldsymbol{AX}=0\}$，$\boldsymbol{A}=(a_{ij})_{m\times n}$，$S=\{(x_1,x_2,\cdots,x_n)\mid\sum\limits_{i=1}^{n}x_i=1,x_i\geqslant 0,i=1,\cdots,n\}$。

我们约定卡玛卡标准型满足下列两个假设条件：

H_1：它的最优值为零，即 $\min\boldsymbol{C}^T\boldsymbol{X}=0,\boldsymbol{X}\in\Omega\cap S$；

H_2：单纯形 S 的中心 $\boldsymbol{X}^0=\frac{1}{n}(1,\cdots,1)^T\in\Omega$。

由于计算比较复杂，卡玛卡在此基础上又给出一种"内点法"。此算法的基本思想是：先假设 LP 问题内存在内点 $\boldsymbol{X}^{(0)}$，并假定问题是有界的。从内点 $\boldsymbol{X}^{(0)}$ 出发，沿可行方向求出使目标函数值上升的后继点，再从得到的内点出发，沿另一个可行方向求出使目标函数上升的内点。重复以上步骤，产生一个由内点组成的序列 $\{\boldsymbol{X}^{(k)}\}$，使得 $\boldsymbol{C}^T\boldsymbol{X}^{(k+1)}>\boldsymbol{C}^T\boldsymbol{X}^k$。当满足终止准则时，则停止迭代，最后得到最优解。这种方法的关键是选择使目标函数值上升的可行方向。其实，在第 13 届国际数学规划会议后，人们受卡玛卡算法的启发发明了一系列的新算法，这些新算法统称为"内点法"。对内点法的研究已形成探索数学规划的一种时代潮流，如我国清华大学应用数学系对内点法作了深入的探索，给出了一种不同于对偶型内点算法的混合算法，此法具有计算简单、实用效果好的优点，在内点法中独树一帜。

4. 鞍面算法

我国科学家尚毅，1981 年在解一种类型的控制问题时，找到了用"鞍面法"来解线性规划的思路。1985 年他在沈阳电子研究所任职时正式提出鞍面法。这种方法主要是按梯度方向逼近"鞍点"的原理来求 LP 的最优解的最新方法。原电子工业部曾组织专家鉴定，认为尚毅的鞍面法是低次多项式算法，与变量维数无关。此后，尚毅按专家们提供的意见，经过两年多的艰苦探索，在大型计算机上做了上万道 LP 的问题，其中包括美国贝尔公司向他提出的几十个难题。他对计算结果深入分析，采用了全新的数学推理思想和计算方法，进一步完善了鞍面法的数学理论，终于解决了鞍面法的关键——"鞍点"向二次极值转换的计算机迭代公式，提高了算法的计算速度和计算精度。他用鞍面法编制的

程序，在计算机上计算有 2 万个变量和 1 万个约束的 LP，只需迭代 300 次左右，而用单纯形法的标准软件解同样规模的 LP，需要迭代 2 万次以上。更重要的是，鞍面法的软件可在个人计算机上解出具有 1 万个以上变量 LP 问题，这是目前国际上任何其他计算机软件做不到的。

解大型复杂线性规划的鞍面法，是我国科学家在国际上继美国单纯形法和卡玛卡算法之后提出的又一独创的新算法。这项成果于 1990 年 7 月正式通过辽宁省科委组织的专家鉴定，被确认为达到国际先进水平。

尚毅的鞍面法将广泛应用于国民经济，特别是用于制定国家或地区国民经济计划，同时也为优化大型工程的设计提供了新的有力工具。例如，1991 年，鞍钢采用尚毅发明的大型线性规划算法及软件，提高经济效益达 1 000 万元以上，使全公司的产品数量、品种和效益同步增长。鞍钢信息管理中心在使用鞍面法解题器的同时，也使用了美国斯坦福大学的软件包。经反复研究比较，无论是从速度还是从精度上，前者皆优于后者。

鞍面法不仅具有重要的学术价值和实用价值，而且标志着我国科学家在国际数学规划领域也拥有一席之地。□

4 无约束非线性规划的数值方法*

在自动控制、科学管理和其他一些科学领域中，很多问题可以归结为线性规划。但是，还有另外一些问题，其目标函数和约束条件却很难用线性函数来表达，我们称这种数学规划为非线性规划。1951 年，库恩(Kuhn)和塔克(Tucker)提出了处理非线性规划问题的完整理论，得到一组不等式约束最优化问题的必要条件，称为库恩—塔克定理。非线性规划到目前为止，还没有适用于各种问题的一般算法，各个方法都有自己特定的适用范围。

优化包括数学规划的全部内容，最简单的优化问题是无约束最优化。对这类问题，已经研究出比较有效的求解方法，因而构成更一般非线性规划解法的出发点。

设目标函数是一个无约束的变量函数 $f(\boldsymbol{X}),\boldsymbol{X}\in R^n$，其中 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$ 是 n 维向量。我们知道，如果目标函数在其定义域内连续且存在一阶偏导数，则目标函数在点 $\boldsymbol{X}^*$ 处达到极值的必要条件是函数 $f(\boldsymbol{X})$ 对各个变量的一阶偏导数在该点等于零。因此，只要找到使 $\nabla f(\boldsymbol{X})=0$ 的点 $\boldsymbol{X}^*$，问题就解决了。但在实际的工程规划中，由于目标函数是高次多元函数，而且求导也存在困难，因而，在一般情况下，关于无约束非线性规划问题的解法常分成两类：解析法和数值计算法。

1. 解析法——间接最优化方法。解析法要首先求出目标函数的一阶、二阶导数，然后求解由目标函数的偏导数所组成的方程组，以便求出稳定点，再用梯度矩阵及海

---

*本文原载于《自动化博览》1997 年第 2 期。

森(Hesse)矩阵对所找到的稳定点进行判断，看它是否为最优点。在目标函数比较简单时，求解上述方程组和利用海森矩阵进行判断并不困难，但当目标函数比较复杂或为非凸性函数时，应用此法会带来麻烦，有时甚至很难求解。

2. 数值计算法——直接最优化方法。直接法，这是一种建立在电算技术基础上的方法，无论在处理方法与概念上与经典方法、理论都不相同。它的特点是：① 是一种直接的数值解法，而不是分析方法；② 按一定的逻辑结构，反复地重复进行同一类型的运算；③ 它是根据目标函数的变化规律，以适当的步长沿着使目标函数值下降的方向逐步向目标函数值的最优点进行探索并逼近，故得出的是近似解，而非精确解，称为逐次逼近法或迭代法。

迭代法的简单过程大致可归纳为以下步骤：① 首先初步选一个尽可能靠近最优点的初始迭代点 $\boldsymbol{X}^{(0)}$，从该点出发按一定规则寻找可行方向和初始步长，向前跨出一步达到 $\boldsymbol{X}^{(1)}$ 点；② 再选择一个新的使函数值下降的方向和适当的步长，从 $\boldsymbol{X}^{(1)}$ 点出发再跨出一步达到 $\boldsymbol{X}^{(2)}$ 点，并依次类推，最终逼近或达到目标函数的最优点。在中间过程中每一步的迭代公式为

$$\boldsymbol{X}^{(k+1)}=\boldsymbol{X}^{(k)}+A^{(k)}S^{(k)},$$

使 $f(\boldsymbol{X}^{(k+1)})<f(\boldsymbol{X}^{(k)})$，$k=0,1,\cdots$，即应使目标函数值一次比一次减小。可见迭代算法有三个要素：初始迭代点 $\boldsymbol{X}^{(0)}$，最优迭代方向 $S^{(k)}$ 及最优步长 $A^{(k)}$。

迭代算法可以有许多种，算法的好坏取决于收敛速度。收敛问题是迭代算法中最重要的理论问题。各种算法按收敛速度可分成三类：① 线性收敛，程序简单，计算方便，但速度较慢，如梯度法；② 二阶收敛，收敛快，但计算量较大，程序复杂，如牛顿法；③ 超线性收敛，介于以上两类之间，属较好一类，如共轭梯度法、变尺度法等。

另外还有一个判别算法收敛好坏的标准，即算法是否具有二次收敛性。所谓二次收敛性是指当一个算法具有对称正定矩阵的二次函数时，可在有限步迭代内获得它的极小点。这种二次收敛性与前述的二阶收敛之间没有必然的联系，更不是一回事。但一般来说，具有二次收敛性的算法往往具有超线性以上的收敛阶数，因而属于比较好的算法。

在迭代过程中，由于真正的最优点在哪里事先并不知道，只能根据相继两次迭代的结果决定何时停止计算。常用的终止计算准则有以下三种：① 根据相继两次迭代的绝对误差；② 根据相继两次迭代的相对误差；③ 根据目标函数梯度的模足够小。

满足上述条件之一时，则认为目标函数 $f(\boldsymbol{X}^{(k+1)})$ 值收敛于函数 $f(\boldsymbol{X})$ 的最小值。这样就求得近似的最优化解

$$\boldsymbol{X}^{*}=\boldsymbol{X}^{(k+1)},\ f(\boldsymbol{X}^{*})=f(\boldsymbol{X}^{(k+1)}),$$

迭代过程可以结束。□

5 约束非线性规划*

一 什么是约束非线性规划

在实际工程和管理科学中有很多问题，其目标函数和(或)约束条件很难用线性函数来表达，如果目标函数或约束条件中有一个或多个是自变量的非线性函数，则称这种数学规划为非线性规划。实际问题中没有约束的优化问题，几乎是没有的，工程设计、运筹学、过程控制、经济学及其他数学领域的许多定量问题都可以表示为非线性规划问题。

解非线性规划要用非线性规划的方法，它比解线性规划困难得多。因为解线性规划有行之有效的通用算法，如单纯形法及修正单纯形法、卡玛卡算法和我国科学家尚毅发明的“鞍面法”等等。但非线性规划，到目前为止，还没有建立起适用于各种问题的一般算法，各个方法都有自己的特定范围，特别是解约束非线性规划比解无约束非线性规划更为困难。

线性规划中，满足约束条件的解称为可行解，所有可行解的集合称为可行域，使目标函数取极值的可行解称为最优解。对于线性规划问题，如果其最优解存在，则其最优解只能在其可行域的边界上达到，特别是在可行域的顶点上达到；而非线性规划的最优解(如果存在)，则可在其

---

*本文原载于《自动化博览》1993年第5期。

可行域的任一点上达到。由于线性规划的目标函数是线性函数，可行域为凸集，因而求出的最优解是全局最优解。非线性规划则不然，有时求出的可行解虽是可行域上的极值点，但并不一定是整个可行域上的全局最优解，只有在一定条件下，局部极大或极小也是全局的极大或极小。

二　约束非线性规划的解法

解约束非线性规划问题的方法，大致可分为两类：

1. 间接法——解析法。这种方法是将约束问题转化为一个或一系列无约束问题，解约束非线性规划问题，当然必须依靠无约束非线性规划问题方面的基本理论。无约束非线性规划方面的理论是十分丰富的，它为约束问题的解决提供了基本条件，特别是一维搜索方面的基本理论和技巧，更是整个优化问题的基础内容，它启发我们将约束问题转化成一个或一系列无约束问题，其基本思想是：根据约束条件，改造目标函数，构造出一个既能表达原目标函数又能包括原约束条件的无约束新目标函数，然后用无约束最优化方法求解，使最优解能足够近似地代替原函数的最优解。求解的过程，实际上是一次次地逼近，用一系列的构造函数去逼近求解，而在最后得到一个解的系列，并且得到一个极限，即最优解。这类方法对等式约束和不等式约束都有效，因为不等式约束可以通过引入松弛变量转化为等式约束。由于这类方法是利用数学分析的方法，根据目标函数（或泛函）与约束函数的变化规律，求出最优解的必要条件：一组含有导数（或变分）的方程式或不等式，得到最优解的候选点（例如平稳点、库恩—塔克点），最后利用充分条件或其他方法从候选点中确定最优解，所以是一种解析方法。

拉格朗日乘子法是一种将约束问题转化为无约束问题的最优化方法，是解决具有附加约束条件的数学问题的工具。此法通常是通过一些称为乘子 λ_i 的待定系数，把原有的约束条件与这些乘子相结合，并添加到原目标函数中去，构成一个无约束的新目标函数（拉格朗日函数）。新目标函数的无约束最优解，就是原约束问题的最优解。解决这一问题的关键是引入系数 λ，它表示在极值点附近目标函数值的变化与约束函数变化的比率关系。称 λ 为拉格朗日乘子，其实质是将原目标函数与约束函数经过一番适当的变换，构造成一个约束条件下的新目标函数，然后采用一系列无约束优化问题的解法去求解有约束条件的优化问题。

拉格朗日乘子法在理论上很重要，然而用拉氏函数求解约束条件的优化问

题仍然是非常麻烦而繁琐的，实际工程上不能直接应用，因而必须寻求更实际的，但又非常可靠的方法。拉氏理论的思路提供了解决实际技术问题的线索和途径。许多约束最优化方法的近代成果，实际上都可看成是它的推广，如罚函数法，其理论系统便是来自拉格朗日理论。罚函数法是一种对实际计算和理论研究都非常有价值的方法，也是目前求解约束问题的有效方法之一。它有多种形式（如内点法、外点法、内外点结合法等），其本质都是通过引入惩罚因子把一个约束问题转换为一个一系列无约束问题求解。

2. 直接法。即直接求解约束问题，其基本思想是直接在可行域内寻找约束最优解，这类方法适用于仅含不等式约束条件的问题。当原问题含有等式约束时，应先用消元法或变数变换把等式消去，只有在这种消去能实现时，才能使用这种方法。

三　用线性规划逐步逼近非线性规划的方法

数学中解决非线性问题的一种基本方法是用一个或一系列线性问题来逼近非线性问题。特别是线性规划已经有了一些比较完善而有效的通用解法（如前所述），可以成功地应用于求解各种线性规划这一事实，启发人们设法把非线性规划问题变换为线性规划问题求解，从而产生了各种线性化方法，如用线性逼近法求解线性的约束条件下的非线性规划和非线性约束条件下的非线性规划。又如可行方向法是通过基本迭代格式求解约束优化问题的一类重要方法，其中关键的一步是通过求解一个线性规划问题来确定下降可行方向，所以实际上也是一种线性化方法。

四　约束非线性规划方法的比较与选择

非线性规划是一种优化方法，求出非线性规划的最优解是优化方法的目标。为达此目标，使用哪些方法才能做到既精确又高效，这本身也是一个优化问题。为了节省时间和寻优费用，各种方法的能力大小、它们的优缺点以及应该选用哪个算法等都是必须考虑的问题。目前主要是通过大量的数值计算来比较各种算法的性能。

前已说过，近二三十年来，约束非线性规划方法有了很大的发展，早期的一些方法，如线性逼近法、投影梯度法、可行方向法以至原来被认为是有效的罚函

数法，现在似乎都有些过时了。而 20 世纪 70 年代提出的一些新方法，例如乘子法、精确罚函数法、广义简约梯度法和二次逼近的方法（其中有代表性的一种是约束变尺度法，即 CVM 法），已逐渐成为重要的优化方法，并且随着相应的计算机实用软件的出现，这些方法已进入工程优化的实际应用阶段。

比较算法好坏的标准主要看有效性和可靠性两个方面：① 算法的有效性主要看计算时间的多少和函数与导数计算的次数。计算结果表明，目前最有效的四种算法依次是：二次逼近法、广义简约梯度法、乘子法、罚函数法，其中乘子法与广义简约梯度法之间相差很小，而与程序处理的好坏有关。② 算法的可靠性主要是看解题成功的比例，这与程序编写技巧有很大关系，因而要结合具体程序来比较。可靠性最好的几个程序为：GRGA、GRGI 和 POWELL 的变尺度法程序 VFOIAD。

到目前为止，普遍认为二次逼近法与广义简约梯度法是解约束非线性问题的最好方法。但是这两种方法的算法与程序都比较复杂，且要求问题中的函数与导数有解析表达式。如果某些工程中的函数没有解析表达式，或者函数的导数相当复杂，求值困难，或者没有上述两种方法的已录制好的、经过考验的、立即可上机运行的程序，需要自己编制和调试软件（对于单纯为了应用，则不是唯一研究算法），在这些情况下，其他各种算法仍然值得采用，它们在不同的场合具有不同的优点。如果求解的是个小型问题，计算精度不要求太高，则使用随机搜索法较为适宜。这种方法的程序编制与调试工作量很小，使用的简便性也较好。对于有明确的函数分析表达的中小型问题，罚函数法较好，其中又以内点法较为有效，其程序简单，编写容易，且灵活性较大。

对于不易求导的问题，可以选用不需要计算梯度的无约束方法；对于函数为高度非线性的问题，可以选用数值稳定性较好的变尺度 BFGS 方法等等。

对于具有线性约束的非线性问题，选用梯度投影法处理效果较好；若是问题的规模较大而且导数容易求得，则选用某些可行方向法可能会获得较好的效果。

对于函数不连续、不可微或虽然可微但导数很难求得的问题，当然只有直接搜索法最有效。这类方法在工程最优化应用中使用得较多，因为其算法的逻辑结构比较直观，容易理解。□

6 求无约束极小值的变尺度法*

在最优化问题中有的不受约束条件限制，称为无约束最优化问题。但大多数最优化问题，往往受到各种因素的制约，称为约束最优化问题。由于约束最优化问题大都可通过不同途径化为无约束问题来求解，所以，无约束最优化方法是解决约束最优化问题的基础。

无约束最优化方法大体上可分为两类：一类是直接法，即直接由目标函数求最优解的方法；另一类是间接法，即通过使用函数的导数求最优解，有时还要根据海森矩阵（由若干个二阶偏导数按一定排列所组成的矩阵）所提供的信息而构成各种方法，这些方法统称为梯度法。一般来说，无约束最优化问题的求解，可通过一系列一维搜索来实现，而搜索方向的选择是解决无约束最优化问题的核心问题。搜索方向的不同选择，形成不同的最优化方法。如最速下降法是选择一个使目标函数值下降得最快的方向，以尽快地达到函数的极小值点。该方法在远离极小值点时收敛速度较快，而当接近极小值点时目标函数下降得很慢。为了克服这一现象，出现了牛顿法。它的最突出的优点是在极小值点附近收敛得很快，但在远离极小值点时很不可靠。牛顿法每次迭代都等价于解一个线性方程组，计算量为向量维数的三次方。当问题的维数增大时，效率急剧降低，而最速下降法只要计算函数的梯度就够了，因此最速下降法与牛顿法各有优缺点。这就启示我们，开始时采用最速下降法，沿负梯度方向搜索；而当试验点离极小

---

*本文原载于《自动化博览》1995 年第 1 期。

点很近时，改用牛顿法，以加快收敛速度。变尺度法正是近20多年来从这种想法出发而构造出来的一种求无约束极小值问题的最有效的方法。

所谓“尺度”，就是空间中对点与点之间的距离的一种度量。例如在 n 维欧氏空间 R^n 中，点 $\boldsymbol{X}(\boldsymbol{X}\in R^n)$ 到原点的距离可以定义为 $\|\boldsymbol{X}\|_2=\sqrt{\boldsymbol{X}^T\boldsymbol{X}}$。这种几何形象的长度，我们称之为欧氏范数或范数2意义下的“尺度”。如果存在一个 $n\times n$ 对称正定矩阵 $\boldsymbol{A}$，$\boldsymbol{X}\in R^n$，那么 $\|\boldsymbol{X}\|_A=\sqrt{\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}}$ 或 $\|\boldsymbol{X}\|_A^2=\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$，其中 $\|\cdot\|_A$ 也是向量的一种范数，即非欧范数，也可以作为对于 R^n 向量($\boldsymbol{X}\in R^n$)进行度量的一种“尺度”，称 $\boldsymbol{A}$ 为尺度矩阵。在欧氏范数下，尺度矩阵 $\boldsymbol{A}$ 变为单位矩阵。

梯度法是在常规的欧氏尺度意义下的最速下降法，而牛顿法不过是在特殊条件下、非常规条件下的最速下降法。

设目标函数为 $F(\boldsymbol{X})$，其中 $\boldsymbol{X}=(x_1,\cdots,x_n)^T$ 为 n 维列向量，符号“T”为向量的转置运算。若用矩阵符号表示沿最速下降方向搜索时，则有迭代公式

$$\boldsymbol{X}^{(k+1)}=\boldsymbol{X}^{(k)}-\lambda_k\boldsymbol{I}g^{(k)},$$

其中 $\boldsymbol{I}$ 为单位矩阵，λ_k 为最优步长，$g^{(k)}$ 为 $F(\boldsymbol{X})$ 在 $\boldsymbol{X}^{(k)}$ 处的梯度方向，称为搜索方向。而改用牛顿方法进行搜索时，则有迭代公式

$$\boldsymbol{X}^{(k+1)}=\boldsymbol{X}^{(k)}-\lambda_k\boldsymbol{A}^{-1}g^{(k)},$$

$\boldsymbol{A}^{-1}$ 为海森矩阵 $\boldsymbol{A}$ 的逆矩阵。

1959年，戴维敦(Davidon)首先提出了变尺度法，后经付莱丘(Fletcher)、鲍威尔(Powell)于1963年进行简化和证明，并以三位学者名字的首字母D、F、P命名为DFP法。

变尺度法的迭代过程，是迭代矩阵 H^k 逐步逼近海森矩阵 $\boldsymbol{A}$ 的逆矩阵 $\boldsymbol{A}^{-1}$ 的过程。从这个意义上说，变尺度法实质上是一种用一阶导数来估计二阶导数的算法。对具有正定矩阵的二次函数来说，DFP法将产生一组 H 共轭的搜索方向，并具备有限步收敛的性质。DFP法成为60年代初期寻找局部极小值的方法中之最有力的通用方法，它在优化算法上有了突破。例如在1962年变尺度法没有得到应用以前，由于原有的各种方法用计算机耗时太多，因此求非线性函数的极小值，一般只能计算10个以下的变量；1962年以后应用了DFP变尺度法，可以在几分钟内计算出有100个变量的函数的极小值。有的问题用变尺度法只要半分钟就可求出最优解，而在以前用其他方法，最快也要30分钟才能求出。

变尺度法问世后，得到了人们广泛的关注，一时成为人们研究的热门课题之一。人们发现变尺度法既有一些好的特性，也有其不足之处。例如60年代

后期，很多人在大量的计算实践中，发现 DFP 法在数值稳定性方面存在一些问题。检查发现产生不稳定的原因是多方面的，例如计算机的舍入误差，目标函数中存在非二次项，一维搜索不够精确等等，从而使搜索方向不是共轭的，甚至不是函数下降的方向。于是，1965 年以后，出现了许多比 DFP 法有更好的数值稳定性的新的变尺度法。这些方法的差异，主要表现在两个方面：其一是迭代矩阵的计算公式；其二是一维搜索的类别和精度。对为数众多的变尺度法，1970 年，Huang 对其做了统一处理，得出了实质上只有 3 个自由参数的统一公式，只要在此公式中对参数赋以不同的值，便可得出很多种不同的变尺度法，而且常见的著名的变尺度法都可以通过 Huang 的公式取不同的参数值而得到。人们把由这个统一公式所产生的变尺度法，称为 Huang 族变尺度法。在理论上，Huang 族变尺度法用精确的一维搜索才能收到较好的收敛效果，也就是达到二次截止的性质。1967 年，Broyden 从另一角度已经导出了一种变尺度算法族。但在 1970 年，Huang 证实了它不过只是被包含在具有二次收敛性的更大的 Huang 族变尺度法的一支，是它的一个子集或子族。1970 年，Broyden、Fletcher、Godfarb、Shanno 等人提出了一个更新的算法，称为 BFGS 变尺度法，它可避免产生奇异，因此 BFGS 法具有更好的稳定性，公认为是最有效的一种算法。不过，此后通过实际计算，又发现单独使用 DFP 或 BFGS 算法有时也发生困难。如单独使用 DFP 法，矩阵 $H^{(k)}$ 有时会变为奇异；如单独使用 BFGS 算法，虽然一般情况下效果很好，但偶尔矩阵 $H^{(k)}$ 会变为无界。因此，Fletcher 于 1972 年提出了把这两种算法组合起来的一种开关算法，虽然也是目前最有效的一种算法，但它的推导过程十分复杂。

总之，DFP 算法提供了一种全新的算法，对于最优化理论的开拓起了不可估量的作用，是最优化理论发展史上的一个里程碑。□

7 多目标最优化问题

一 什么是多目标最优化问题

我们知道，线性规划是对线性系统进行优化和决策的有效工具。它的基本特点是：在一定约束条件下，实现目标函数的极小值或极大值。但由于它的目标函数是单一的，即单目标函数，故难以适应现代化生产与管理中错综复杂的多目标的实际需要，因为在实际应用中，更多地会遇到多目标优化问题。其次，线性规划有它的局限性，就是它的最终结果只有两种情况：要么有最优解，要么无可行解，没有给决策者留下分析和选择的余地，从而也限制了线性规划的应用范围。目标规划就是针对线性规划的这些局限性而发展起来的一种优化方法，它是解决约束条件相互矛盾、目标之间重要程度不同的多目标规划。

原来，对一项复杂的工作进行决策时，往往有许多可供选择的方案（称为候选方案），人们为了取得的效果，总是要设法按某种准则（目标或指标），从众多候选方案中选出最优或接近最优的方案。若所考虑的问题只有一个目标作为选择好坏的标准，而方案又不多时，利用简单的方案比较法，即可选出其优者。这类问题就是通常所说的单目标最优化问题。但是，实践中出现的许多问题往往不是一个目标所能表达的，比如对一架轧机来说，如果要求轧机重量最轻的同时，又希望刚度最大（或者说轧机变形最小），这就是一个双目标最优化问题。在设计一个新的工

艺过程时，往往希望优质、高产、低消耗，这是要求三个同时达到最优的目标。又如在确定一个系统设计方案时，往往考虑到高可靠性、高精度、周期短、成本低、维护方便等等。这类在给定的某些条件限制下，同时要求多个目标尽可能好的最优化问题，就是所谓多目标最优化问题。

研究多目标最优化问题的学科称为多目标最优化或多目标规划，它是运筹学，也是数学规划的一个重要分支。用数学的语言来说，多目标最优化研究的对象是多于一个的数值目标函数在给定的区域上的最优化（极小化或极大化）问题。由于多个数值目标可用一个向量目标表示，因此，多目标最优化问题有时也叫向量极值问题。不失一般性，多目标函数最优化问题的数学模型可写为以下形式：

$$\begin{cases} \min F(\boldsymbol{X}), F(\boldsymbol{X})=(F_1(\boldsymbol{X}), F_2(\boldsymbol{X}), \cdots, F_n(\boldsymbol{X}))^T, \boldsymbol{X}\in R^n, \\ g_i(\boldsymbol{X})\leqslant 0 (\text{或 } g_i(\boldsymbol{X})\geqslant 0), i=1,\cdots,l, \\ h_j(\boldsymbol{X})=0, j=1,\cdots,m, \end{cases} \tag{1}$$

式中 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$ 称为决策变量，$F(\boldsymbol{X})$称为向量目标函数，它们具有不同的定义和完全独立的关系。在几何上，决策变量一组取定值，对应于 n 维欧氏空间中的一个点。满足约束条件的点称为可行点，所有可行点组成的集合称为可行域，记为 S。上式也可简记为

$$\text{V-}\min_{\boldsymbol{X}\in S} F(\boldsymbol{X}) \quad (\text{VMP})。 \tag{1'}$$

上述的 VMP 为向量数学规划(vector mathematical programming)的缩写。VMP 中的 V-min 表示向量极小化，即向量目标函数 $F(\boldsymbol{X})=(F_1(\boldsymbol{X}),\cdots,F_n(\boldsymbol{X}))^T$的各个目标函数被同等的极小化的意思，$\boldsymbol{X}\in S$ 表示决策变量满足所有的约束条件。

特别是，当各个目标函数以及约束条件都是决策变量的线性函数时，则称相应的多目标最优化为多目标线性规划；若进一步的限制决策变量取整数值，则相应的问题叫做多目标整数规划；当各个目标函数、约束条件有一为非线性函数时，则相应的问题叫做多目标非线性规划。

二　多目标最优化问题的解法简述

多目标问题与单目标问题的一个重要区别是，单目标问题的任意两个可行解总是可以比较其优劣的，但对多目标问题就不尽如此，它的两个可行解不一定能进行优劣性的比较。因为各个分目标有些是可以相互结合、统一考虑，予

以平衡或协调的，但还有一些往往是互相影响、互相制约，有的甚至是互相矛盾的，不能同时达到最优解，甚至某一可行点对一个分目标函数是最优点，对另一个分目标函数却是劣点。因此，对于多目标最优化就成为一个复杂而困难的问题。但由于多目标优化的理论与方法，在诸如经济学、决策学、能源开发、工程设计、控制论以及军事等众多领域有着大量的应用，对策论、系统工程等学科也常常涉及到多目标优化问题，因此很早就开始了对它进行研究。早在1896年，经济学家波雷特(Pareto)首先在经济平衡的研究中提出了多目标最优化问题，引进了被称为Pareto最优的概念。1951年，数理经济学家库朴曼斯(Koopmans)从生产和分配效率分析中考虑了多目标最优化问题，引入了有效解的定义并得到某些结果，为多目标最优化学科奠定了初步的基础。20世纪60年代以来，人们设计了求解多目标最优化的方法，并运用它们求解各种实际问题。但由于多目标的复杂性，要找到一个方案比其他方案都好的可能性是很小的，绝对的最优解是没有的，但是“次优”、“相对的优”则是可达的。正如1978年诺贝尔经济学奖获得者西蒙(Simon)所说：在现实世界里，客观合理的举动或者接近这种客观的合理性是难以达到的，所以他提出“令人满意”(satisfying)一词代替“最优化”，似乎更合适。事实上，人们正是按这种原则行事的。比如一个人到商店买东西，总是希望买一件最好的，但实际上他只能买一件他认为满意的，因为他不可能挑遍商店里的全部他所要买的那种商品后再买。现在一个为大多数人所能接受的办法，是在各目标的最优化之间进行协调，相互做出“让步”，以便取得整体最优方案。一类最基本的方法是所谓评价函数方法。评价函数，就是用来评价目标优劣的函数。

求解多目标优化模型(1)的一个重要而基本的途径，是根据问题的特点和决策者的意图，构造一个把n个目标转化为一个数值目标的评价函数$w(F)=w(F_1,F_2,\cdots,F_i)$，通过它对$n$个目标$F_i(i=1,\cdots,n)$的评价，把求解多目标优化问题归结为求解与之相关的单目标(数值)极小化问题

$$\min_{x\in S} w(F)。$$

这种借助于构造评价函数把求解多目标最优化问题归结为求解单目标最优化问题的方法，称为评价函数法。由于可用不同的方法构造评价函数，故有不同的评价函数法。一个最简单也是最基本的评价函数法就是线性加权和法。

1. 线性加权和法。这个方法的指导思想是，根据各个分目标函数在问题中的重要程度，分别赋予它们一个数，并把这个数对应地作为各目标的系数，然后把这些带系数的目标相加来构造评价函数。极小化由该评价函数所构成的数值函数，其最优解即作为原多目标函数极小化问题的解。

具体地说，对于多目标优化模型(1)，设给定一组与各目标 F_i 对应的非负数 $w_i=(1,2,\cdots,n)$：

$$F_1,F_2,\cdots,F_n;$$
$$w_1,w_2,\cdots,w_n。$$

其中各 w_i 的大小代表相应目标 F_i 在模型(1)或(1')中的重要程度(w_i 越大表示 F_i 在(1)中越重要；反之，越小的 w_i 表示对应的 F_i 越不重要)及其量级和量纲。现在把各个目标带上相应的系数得到 $w_iF_i(i=1,2,\cdots,n)$，再对它们求和，构造出一个称为评价函数的新目标函数

$$M(F)=\sum_{i=1}^{n}w_iF_i。\tag{2}$$

通过这个评价函数，便可以把几个目标函数综合起来转化为一个数值函数

$$M(F(\boldsymbol{X}))=\sum_{i=1}^{n}w_iF_i(\boldsymbol{X})。\tag{3}$$

然后极小化上述的数值函数(3)，即求解

$$\min M(F(\boldsymbol{X}))=\min_{\boldsymbol{X}\in S}\sum_{i=1}^{n}w_iF_i(\boldsymbol{X})。\tag{4}$$

其最优解便是按各目标重要程度的意义下使各目标尽可能小的解。通常，我们把评价函数(2)中反映重要程度的各个 $w_i\geqslant 0(i=1,2,\cdots,n)$ 叫做对应项的权系数，或简称权因子。一般还要求 $\sum_{i=1}^{n}w_i=1$。由于这个方法的主要特点是对各目标“加权”之后，以其线性和作为评价函数，故叫做线性加权和法。

这个方法的关键在于如何找到合理的加权因子 w_i，以真正反映出各个分目标在整个多目标问题中相对重要程度及其量极和量纲上的差异。虽然人们已提出各种确定权系数的方法，但迄今尚没有一个普遍准则。常用的确定权因子的方法是“老手法”，这是一种凭借经验评估并结合统计处理来确定权系数的办法。首先是邀请一些与所研究的问题有关的专家，请他们各自独立地对问题中要确定重要程度的几个事项给出相应的权系数，并填写到一定的权系数表格上，汇总后再经充分讨论和交换意见，消除误解(但不交流各“老手”所提的权系数方案)，然后，让他们重新调整系数方案，并将它们再列入权系数方案表。重复上述过程，最后得到一组满意的权系数均值作为问题的权系数。此法简单易行，有实用价值，但邀请的“老手”人数一般不宜太少。

2. 理想点法。假定有 n 个分目标 $F_i(\boldsymbol{X})(i=1,2,\cdots,n)$，并假定希望越小越好，而且每个目标各自的最优值存在，先求出各个单目标的最优值，记为

$$\min F_i(\boldsymbol{X})=F_i^*,i=1,2,\cdots,n,$$
$$F^*=(F_1^*,F_2^*,\cdots,F_n^*)^T。$$

如果有某一个方案使各个目标同时达到最优，那当然是最为理想的，但在一般情况下，这种可能性是极少的。因此，我们称 $F_i^* = (F_1^*, F_2^*, \cdots, F_n^*)^T$ 为理想点。虽然理想点实际上不可能达到，但我们希望找出尽可能接近理想点的方案。于是，我们构造新的评价函数

$$U(\boldsymbol{X}) = \sqrt{\sum_{i=1}^{n} (F_i(\boldsymbol{X}) - F_i^*)^2} \ 。$$

我们把对应于 $\min U(\boldsymbol{X})$ 的解作为原多目标最优化问题的满意解。

由前苏联学者沙罗卡瓦特于 1975 年首先提出的这个方法是很直观的，但我们在应用这个方法时会遇到如下的问题，要求出理想点本身往往是一件复杂的工作。有时，有的问题的理想点根本不存在；按理想点求得的解有时不一定能满足工程设计的要求（假定满足要求的解存在）。针对这些问题，有人提出了如下的虚拟目标法。

3. 虚拟目标法。某些多目标优化问题，往往由于工程实际的需要，要求所选的方案的各项目标必须满足预定的要求，并希望在满足这一预定要求的条件下，再找出尽可能好的方案。例如设计一个导弹，既要其射程远，又要消耗燃料少，还要命中率高等；选择工厂的厂址，除了要考虑运费、造价、燃料供应费等经济指标外，还要考虑其对环境的污染等社会因素。只有对各种因素的指标进行综合考虑后，才能作出合理的决策。

假定有 n 个目标如(1)所示，预定的目标值为 $F_i^0, i=1,2,\cdots,n$，每个目标都要求最好。现在虚拟目标

$$F^0 = (F_1^*, F_2^*, \cdots, F_n^*)^T,$$

其中 $F_i^0 = F_i^0 + P_i > 0$，P_i 的取值由决策者根据自己的愿望或对目标的偏好决断而定，相对重要的目标判断值要大一些。做出新的评价函数为

$$U(\boldsymbol{X}) = \sqrt{\sum_{i=1}^{n} \left(\frac{F_i(\boldsymbol{X}) - F_i^0}{F_i^0} \right)^2} \ 。$$

我们把对应于 $\min U(\boldsymbol{X})$ 的解作为原多目标问题的满意解。

通过评价函数法来解多目标优化问题，虽是一类基本的方法，但是在实际应用中对一个复杂的多目标最优化问题，要根据它的特点构造出一个恰当的评价函数的表达式却是非常困难的。为此，人们提出了其他一些方法，如目标规划法、交互规划法、混合选优法、分层求解法等等，无非也都是照顾几个目标均能处在较合理的范围内，以各分目标为单独目标去解决，然后在确定最优解时适当照顾几个分目标可能满足的程度。例如目标规划是多目标最优化研究中一类十分实用的模型，由于这类方法在处理问题时具有比较灵活和简便的特点，因而在实际应用中为人们所乐于采用。更为重要的是，线性目标规划的算

法和软件已获得了成功的开发。

自 20 世纪 70 年代以来，我国对多目标优化的理论和应用研究日渐深入，其应用范围日益扩大，解决问题的规模也愈加大型化。多目标最优化作为进行重大决策和解决实际问题的强有力的手段和有效的工具，必将在我国现代化建设中发挥重要的作用。□

8 动态规划*

动态规划是解决多阶段(步)过程最优化的一种数学方法。所谓多阶段过程(步)是指一类过程由于它的特殊性，可以将过程按时间顺序分为若干个阶段，而在每一个阶段都需要决策(控制)，以使整个过程取得最优的效果。这一类问题属于规划论问题的范畴，但是在这个问题中，时间是一个很重要的因素，在各阶段所采取的最优策略与时间有密切关系。在各个时间阶段，采用的不同策略是随时间而变动的，这就有“动态”的含义，它是在时间的推移过程中，要在每一个时间段选择最恰当的策略，以期在整体上达到最优。因此在数学上就把这种处理方法称为“动态规划”，这也就是“动态”一词的来由。例如，我们从某发射站发射一枚导弹，其目的是要在给定时间内击中一移动目标(例如敌人的飞机或其他飞行器之类)。假设导弹的速度是能追上此目标的，问题是如何根据目标的行动来每隔一个时间段决定导弹的方向与速度，使之在给定的时间内能击中此目标，这就是一个多阶段的决策(控制)过程问题。但如果是一个与时间无关的静态问题，这时只要在静态模型中人为地引进“时间”因素，并把它作为多阶段的动态模型来考虑，即按某个参变量把原问题分解成一个序贯的多阶段决策问题，那么也可以用动态规划的方法来对其进行求解。

1951 年，美国数学家贝尔曼等人根据一类多阶段(步)决策问题的特性提出了解决这类问题的“最优化(性)原

---

* 本文原载于《自动化博览》1992 年第 2 期。

理”，并研究了许多实际问题，从而创立了动态规划，以解决空间技术中的制导与控制问题。1957 年出版的《动态规划》，系统地论述了这种最优化的数学方法，这是关于动态规划的第一本专著，也是动态规划的奠基之作，它的出版不仅标志着动态规划的正式诞生，并且也把动态规划的研究推向一个新高峰。此后，动态规划被广泛应用于工程技术、经济等部门的最优决策。近年来，在最优控制特别是在随机最优控制领域，也找到了它的应用。

我们知道，20 世纪 50 年代前，处理最优化问题的主要方法是古典的微分法和变分法，两者都可用以求目标函数的极值问题，但是变分法在被应用于处理函数的极值问题时，所求函数的最优性是用一组微分方程来表示的，而这些微分方程的求解一般是比较困难的，特别是在附有两端边界条件时更是这样。动态规划则是用完全不同的一种全新思想方法，提出了以动态规划为基础的、著名的“最优化(性)原理”。该原理为：作为整个过程的最优决策(控制)具有这样的性质，即无论过去的状态和决策如何，相对于前面的决策(控制)所形成的状态而言，余下的诸策略必须构成最优策略。这个原理可以简单地叙述为：每个最优策略，只能由最优子策略构成；或者说，最优策略的一部分也是最优策略。如图 1，若 $SDGE$ 是从 S 到 E 的最优(短)路线，则 DGE 也是从 D 到 E 的最优(短)路线。如果不是这样，则从 D 到 E 必有另一条距离更短的路线，例如 $SDFE$ 存在，这条路线比原来的路线 $SDGE$ 更短，但这与原假设 $SDGE$ 最优(短)是矛盾的。根据这个简单而又直观的最优化原理，可以导出一个函数方程，称为动态规划的基本方程。在最优策略存在的情况下，它是求解最优策略的基本工具。实际上，所有动态规划的函数方程，都是根据这个最优化原理推导出来的。

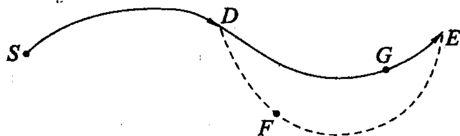


图 1

动态规划求解最优化问题的基本思想方法是：设法把含有许多决策变量的问题分解成一系列前后相关的 n 个阶段(步)，由这 n 个阶段(步)组成几个子问题，每个子问题中仅含有一个或很少几个需要决策的变量，然后顺序地利用已解出的前一子问题的结果，逐一地求解后一子问题，最后一个子问题的解就是所求的原问题的解。只要这种分解是可能的，通常可以大量地减少计算的时间。

这样，利用最优化原理，可以把多阶段(步)决策问题的求解过程看成是一个连续的递推过程，由后向前逐步推算。在求解时，各状态前面的状态和决策对后面的问题来说，只不过是相当于初始条件而已，并不影响后面过程的最优化决策。下面以著名的"最短路线"为例来说明动态规划的基本思想方法。

如图 2 所示，我们从 S 点出发到目的地 F，希望找出一条最短路线。图中相邻两点连线上的数字表示这两点间的距离。把 S 到 F 的全过程分成四个阶段：由 S 到 $B(B_1, B_2)$ 为第一阶段，由 B 到 $C(C_1, C_2)$ 为第二阶段，由 C 到 $D(D_1, D_2)$ 为第三阶段，由 D 到 F 为第四阶段。

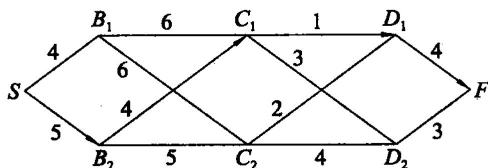


图 2

求出 S 到 F 最短路线的直接方法称为穷举法，即把所有的路线都列出来，然后逐个计算、比较，挑选出其中距离最小者，即为所求最短路线。例如路线 $S\to B_1\to C_1\to D_1\to F$，通过三次加法可算出 4+6+1+4=15。图 2 中从 S 到 F 共有 8 条路线，通过 24 次加法可算得最短的距离为 14，从而得相应的最短路线为 $S\to B_2\to C_1\to D_1\to F$。但当段数很大、各阶段的允许决策很多时，计算量会大得惊人，即使高速电子计算机也力不从心。但若用动态规划法来求，从最后一段开始，由后往前逐段推算，求出各点到终点的最短路线，最后求出全程最短路线，这就简单得多。就本例而言，它的计算顺序就是从最后的第四阶段开始，从 F 到 S 逆序逐个阶段往前计算：

① 第四阶段的计算为 $f_4(D_1)=4$，相应的路线为 $D_1\to F$。又 $f_4(D_2)=3$，相应的路线为 $D_2\to F$。

② 第三阶段的计算为：

$$f_3(C_1)=\min\begin{Bmatrix} d(C_1,D_1)+f_4(D_1) \\ d(C_1,D_2)+f_4(D_2) \end{Bmatrix}=\min\begin{Bmatrix} 1+4 \\ 3+3 \end{Bmatrix}=5,$$

$d(C_1, D_1)$表示 C_1 与 D_1 间的距离。择其较短距离 5，对应的路线为 $C_1\to D_1\to F$。

$$f_3(C_2)=\min\begin{Bmatrix} d(C_2,D_1)+f_4(D_1) \\ d(C_2,D_2)+f_4(D_2) \end{Bmatrix}=\min\begin{Bmatrix} 2+4 \\ 4+3 \end{Bmatrix}=6,$$

择其较短距离 6，对应的路线为 $C_2\to D_1\to F$。

③ 第二阶段的计算为：

$$f_2(B_1)=\min\begin{Bmatrix}d(B_1,C_1)+f_3(C_1)\\d(B_1,C_2)+f_3(C_2)\end{Bmatrix}=\min\begin{Bmatrix}6+5\\6+6\end{Bmatrix}=11,$$

择其较短距离 11，相应的路线为 $B_1\to C_1\to D_1\to F$。

$$f_2(B_2)=\min\begin{Bmatrix}d(B_2,C_1)+f_3(C_1)\\d(B_2,C_2)+f_3(C_2)\end{Bmatrix}=\min\begin{Bmatrix}4+5\\5+6\end{Bmatrix}=9,$$

择其较短距离 9，对应的路线为 $B_2\to C_1\to D_1\to F$。

④ 第一阶段的计算为：

$$f_1(S)=\min\begin{Bmatrix}d(S,B_1)+f_2(B_1)\\d(S,B_2)+f_2(B_2)\end{Bmatrix}=\min\begin{Bmatrix}4+11\\5+9\end{Bmatrix}=14。$$

以上各方程都是动态规划的函数方程，它给出了各步之间的递推关系。根据最优化原理，最短路线为 $S\to B_2\to C_1\to D_1\to F$，最短距离为 14。

于是，由最优化原理可知，从 S 到 F 的最短路线是 $S\to B_2\to C_1\to D_1\to F$，即为图上箭头所示线段。

从上面计算可以看出，动态规划的特点是：一个过程分成几个阶段（步），从后往前反向计算，每一阶段要作出决策，后一步计算使用前一步的计算结果。动态规划是一种迭代法。

动态规划具有一般方法所不具备的优点：① 把一个复杂的问题分解成 n 个简单的子问题来求解，使计算量大为减少。如上面所说的最短路线问题，若用穷举法，要计算 8 条路线，每条路线要用 3 次加法，整个过程要用 24 次，而动态规划总共才用 10 次加法。当阶段数较大且允许决策很多时，这个优点更为突出。此外，动态规划用的是迭代法，所以更便于用计算机计算。② 动态规划的计算结果也十分丰富，不仅能求出从初始状态开始到最终时刻的 n 个阶段的最优解，而且能算出从每个阶段的开始状态到最终时刻的各子问题的最优解，这样的一族信息对于实际工程的分析与决策是十分有用的。

虽然动态规划能利用计算机得出数值解，然而，动态规划并不是由一组规则组成的具体算法，所以在用计算机求解具体问题时，需要按照这种策略，结合实际问题的性质编出具体的计算机解题算法。

决策过程可分为离散确定性、连续确定性、离散随机性、连续随机性四类，从而动态规划方法也有相应的四类。当多阶段决策的分段参量取离散数值时，或者说当过程是由有限段或可数段组成时，则称此过程为多阶段离散决策过程，对应的动态规划为离散变量动态规划；当分段参量是连续参量时，例如当参量为时间且需要在某一段时间内的任意时刻取值作决策时，则称此过程为连续

决策过程，这时的动态规划为连续变量动态规划。若按决策结果是确定的还是按概率分布规律而概率地确定的，则决策过程又可分为确定性和随机性两类，从而又有确定性和随机性动态规划，如上面所说的“最短路线问题”就是确定性的多阶段离散变量动态规划。

在现代控制理论上，动态规划的基本思想是把一个多步最优控制问题化成为多个一步最优控制问题，它的适用范围比较广泛，可用于多输入多输出系统、线性和非线性系统。此外，用动态规划研究控制理论之所以重要，原因在于：① 可以得出离散时间系统的理论结果，线性二次控制问题（如线性多变量系统的数字控制）的容易处理的解是它的特例；② 动态规划的连续形式给出了它与古典变分法的联系，这实际上也是研究最优控制的一种可供选择的方法。

动态规划也存在一些缺点，如当系统的状态变量维数较高时，计算最优控制所需计算机的存储量和计算量都会较大，特别是所需计算机的存储量，常常超过其所能容纳的程度。所以对于高维数（一般指超过六维）的系统，直接应用动态规划，在计算机上还存在一定困难，贝尔曼称此为“维数灾”。为了克服“维数灾”的困难，又出现了一些改进算法，如微分动态规划是一种利用逐次逼近方法去求动态规划最优方程的解的算法，还有“降维法”等，限于篇幅，不作论述。□

9 最优估计理论*

在现实的通讯和控制系统中，由于系统自身及对其观察的测量装置会受到外界的随机干扰(亦称噪声)，致使系统的状态方程和量测方程中都含有随机干扰项。这样，在研究随机过程的最佳控制时，首先遇到的一个重要课题就是滤波问题。通过对一系列带有误差的实际测量数据的处理，得出所期望的数据的估计值，这种数据处理方法叫做滤波方法。滤波与预测问题密切相关。滤波问题是根据到现时刻为止所获得的量测值，对状态作出不同的估计：估计未来某时刻的状态，称为预测；估计现时刻的状态，称为滤波；估计过去某时刻的状态，称为平滑。滤波、预测与平滑对于通讯与控制等信息传递与处理系统来说，都有很广泛的实用意义。

从数学的观点来说，滤波理论是统计学中估计理论近代发展的一个重要分支。估计可分参数估计和状态估计。在做实验时，常把实验结果用曲线的形式表示，需要根据观测数据来估计描述该曲线的方程中的某些参数，这一过程就叫做参数估计。在飞行器导航中，要从带有随机干扰的观测数据中，估计出飞行器的位置、速度和加速度等运动状态变量，这就是状态变量的估计问题。不消说，人们总是希望被估计出来的参数或状态越接近实际值越好，即估计是最优的。所谓最优估计，是指在某一确定的准则条件下，对某种统计意义来说，估计达到最优。由于衡量估计的准则可能是各式各样的，所以最优估计不是唯一的，

*本文原载于《自动化博览》1996年第2期。

它随着标准不同而不同。在估计时要恰当地选择衡量估计的标准。

估计理论一般是用概率统计的工具。无论是控制论还是信息论都需要研究估计理论，特别是估计理论与最优控制更有密切的关系，它成功地把统计的观点引入到通讯理论与控制理论里来，促进了通讯与控制学科的发展。

为了正确地解决参数估计与状态估计问题，首先要研究估计的方法。目前估计有三种方法：

1. 最小二乘法。作为参数估计的开端，早在1795年，德国数学家高斯在他的《天体运动理论》一书中就提出了最小二乘法，也称为最小二乘原理。在参数估计中，经常采用的最小二乘准则是要估计误差平方和为最小。如果对于一个量以相同的精度进行 n 次测量所得的值分别为 $m_1, m_2, \cdots, m_n$，则此时最好的近似值应使得差 $m-m_i$ 的平方和 $\sum_{i=1}^{n}(m-m_i)^2$ 为最小。高斯在200多年前用这个方法解决了天文学上的定轨问题。随后有许多科学家讨论了最小二乘法的各种形式（如一般最小二乘法、加权最小二乘法、递推最小二乘法等），遂使最小二乘法成为通常数据处理中用得最多的一种参数估计方法。如果在数据处理中完全不掌握随机信号的统计特性，一般就用最小二乘法来进行估值。但是最小二乘法也有某些缺点，首先它没有考虑到被估参数和观测数据的统计特性，其次是采用最小二乘法需要进行大量的测量。因此，最小二乘法不是一种最优估计方法。另一方面，由于最小二乘法在计算上相对来说比较简单，因此它也是用得最广泛的估计方法。以下所谈的卡尔曼滤波本质上是高斯原始最小二乘法的一个递推解。

2. 维纳滤波。在第二次世界大战中，由于雷达的发明以及用于防空火力控制的任务，美国数学家维纳为了对控制系统的有关参数进行估计，从电滤波器中的噪声与消息问题出发，提出了在频域中设计最优滤波器的方法，建立了著名的维纳滤波理论和信号预测理论。维纳滤波器是在把测量信号和干扰信号都能表示成有理谱密度的情况下，找出最优滤波器，使得实际输出与希望输出之间的均方误差为最小。维纳之所以能够将统计的观点引入到通讯与控制中来，是因为他证明了在一定条件下，时间序列的总体平均可以用时间平均来代替。如果平稳过程（是其统计特性不随时间而变的随机过程）对任一函数的有限时间平均或有限样本（以概率为1）收敛于总体平均，则称这个过程为各态历经过程或称为遍历过程。对于平稳的遍历过程，其自相关函数为其功率谱密度的傅里叶反变换，这就是著名的维纳—辛钦定理（遍历定理）。维纳运用这个定理得到了一个重要结论：在一定条件下，处在统计平衡的时间序列的时间平均等于相平均。有了这个前提，就可以从时间序列的过去数据推知未来和预测未

来,维纳正是基于这点提出了他的著名的滤波和预测理论的。如前所述,滤波问题就是尽可能地恢复一个被噪声干扰了的信息流问题。实质上,就是预测一个被噪声干扰了的时间序列的问题,因此,滤波问题也可以视为一个预测问题。从数学上讲,预测就是从一个时间序列的过去的数据来估算整个序列的统计参数,这种估算得出的平均值与客观实际是有一定差距的,最佳预测应使这种误差为最小。维纳滤波所解决的就是在最小均方误差准则下的滤波问题。

设计维纳滤波器的过程就是寻求最佳滤波器冲击响应的明确表达式,其实质就是求解维纳—霍夫方程。维纳运用概率论中的协方差函数描述了滤波器中的信号与噪声的统计特性,在滤波器的随机输入(包括信号与噪声)条件下,可以通过维纳—霍夫方程求解,这个方程是积分方程,其解是维纳滤波器的最优冲击响应。维纳—霍夫方程只有在平稳条件和有理谱密度的情况下才能求出它的解。事实上,由于这个方程的条件较严格,所以求解存在很大困难。在一般情况下,这个方程没有解析解;即使在少数情况能得到解析解,而在工程上也往往难以实现,而且维纳滤波是在频域范畴对线性定常系统而作的,对多变量的时变系统或非线性系统都难以应用。

同一时期,前苏联学者柯尔莫格洛夫也论述了离散时间的滤波问题,提出并初次解决了离散平稳随机序列的预测和外推问题,与维纳滤波一起统称为维纳滤波。

由于军事上的原因,维纳滤波在第二次世界大战期间得到了广泛的应用,但由于它存在着以上那些弱点,限制了它的应用范围与价值。到 20 世纪 50 年代中期,空间技术日益发展,维纳滤波的理论与方法也就越来越不适应实际的需要,于是卡尔曼滤波应运而生。

3. 卡尔曼滤波。维纳滤波的困难问题,首先在 50 年代中期确定卫星轨道的问题上遇到。1958 年,美国斯韦尔林(Swerling)首先提出了处理这个问题的递推算法,并立即被承认和应用。1960 年,美国数学家卡尔曼运用递推算法,成功地把状态变量法引入到滤波中来。用消息与干扰的状态模型代替了通常用来描述它们的协方差函数,将状态空间描述与离散时间更新联系起来,适于计算机直接进行运算,而不是去寻求滤波器冲击响应的明确公式。这种算法得出的是表征状态估计值及其均方误差的微分方程,给出的是递推算法,这就是建立在状态空间时域公式基础上的最优递推滤波方法,这种方法就是现在熟知的卡尔曼滤波器。

卡尔曼滤波有两个方面的明显特点:一是运用递推算法,一是将状态变量引入到滤波理论中,使之能处理多变量系统的滤波问题。

与整段滤波的维纳滤波不同,卡尔曼滤波是递推滤波,不是寻求解析解,而

是在数字机上得出递推的数字解。

当新的数据测得之后，根据新的数据和前一时刻的诸量的估值，借助于系统本身的状态转移方程，按照一套递推公式（差分方程或微分方程），即可算出新的诸量的估值，因此大大减少了存储量和计算量。卡尔曼滤波一经出现，就在多方面得到广泛的应用，例如在测轨问题和惯性导航等方面都应用卡尔曼滤波方法估计出这个飞行器每时每刻的状态变量，如飞行器的位置、速度、加速度及阻力系统等物理量，以便对飞行器进行导航、制导和拦截。此外还可应用于潜艇导航、火力控制、飞机导航、工业生产的最优控制和气象预报等。

维纳滤波的关键是求解维纳—霍夫方程，以期得到最佳线性滤波器的冲击响应，但求解维纳—霍夫方程只有在平稳条件和有理谱密度情况下才是可行的。卡尔曼避开了解维纳—霍夫方程的困难，不去给出消息模型的协方差信息，而是给出类似于动态线性系统的状态模型，即以消息与噪声的状态模型代替维纳滤波中的消息与噪声的协方差函数，这或许是卡尔曼滤波成功的一个重要原因。

卡尔曼滤波与维纳滤波两者的基本区别在于，卡尔曼滤波是用时域及状态变量的方法，而维纳滤波是用频域及传递函数的方法。

目前工程上使用的方法都是近似的，因此得到的结果还不是严格的最优，甚至不能从根本上来评价各种方法的优劣，因为还没有得到衡量这种最优的准则，因而这是值得深入研究的一个领域。□

10 七桥问题和四色问题*

数学史上有许多发人深思、饶有趣味的著名难题，虽已时过境迁，但其价值仍存。七桥问题及四色问题便是其中的两例。

一　七桥问题

在18世纪时，东普鲁士首府哥尼斯堡（现为前苏联的加里宁格勒）有普雷格尔河横贯城区，有七座桥将河两岸及河中两个小岛连接起来（见图1）。当时，那里的居民有一个饶有趣味的数学游戏：一个人从陆地出发，不许重复地走遍这七座桥，回到原出发地，这就是著名的“七桥问题”。不知道有多少人想试试看，但多少年过去了，并没有得到答案。

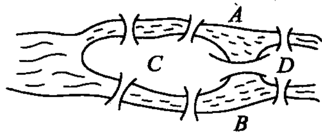


图 1

这个问题传到了当时在圣彼得堡的鼎鼎大名的瑞士数学家欧拉的耳朵里，引起了他的深思。经过研究，他首先简化了这个问题的表示方法，用点代表陆地，用弧或线段代表桥，从而得到一个由四个点和七条弧或线段组成的图形

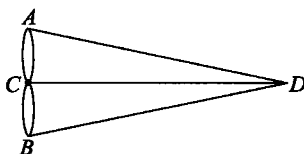


图 2

* 本文原载于《自动化博览》1990年第2期。

（见图 2）。显然，岛的大小和形状以及桥的长度都无关紧要，桥与桥的连接形式才是本质的。这样，对于七桥问题的研究，就变成对四个点和七条弧或线段组成的图形的研究。

欧拉认为，对具体的七桥问题，先仔细地列举出所有可能通过桥的全部路线，然后逐个检验，去掉那些在任一个桥上通过两次或两次以上的路线，就定会得到问题的答案。但这种方法可能的组合数很多，因而是一个繁难而又费力的方法；而在另外一些问题中，如果存在更多的桥，就可能使此法毫无实用价值。欧拉经过探索，把七桥问题归结成"能否一笔画出这个图"的问题。即用铅笔连续不断地画出这个图，在每一条弧或线段上只准画一次，即一笔画问题。如果对画图的过程稍加分析，就不难发现这实际上是把点和线相间排成一串。图中每一顶点如果有一条"进来"的线，就必须有一条"出去"的线。所以除了起点和终点外，图中每一点应与偶数条线相连（即为偶点）。如果起点和终点重合，则这一点也应是偶点。"一笔画"中，奇点（即与奇数条线相连的点）的个数不能多于两个。于是欧拉得出以下结论：

如果某个图满足下面的两个条件，则经过图中每条线一次且仅一次的路线必存在：

（1）从图中任意点出发，通过一些线一定能达到其他任意点；

（2）图中每一点（可能有两点例外）必须是偶点。

所谓图，是由若干点（称为顶点）的非空有限集和连接这些点中的某些"点对"的连线（称为边）所组成的图形。一个图是连通的是指这个图中任意两点之间至少有一条链，否则称为不连通的。

欧拉对以上结论给出了严格的数学证明。为了纪念他的这一贡献，凡从图中任一点出发，最后能回到原来出发点的图，即能一笔画出的图称为欧拉图。他证明了对于图 2 一笔画出是不可能的，即七桥问题没有解，等价于图 2 中的图不能以某种方式走遍（因为图 2 奇点个数是 4）。故断言：要想不重复地一次走遍这七座小桥是不可能的。

许多人为其绞尽脑汁而百思不解的七桥问题，欧拉竟如此简单而巧妙地解决了，从而赢得了人们的高度赞赏和钦佩。七桥问题如上所说，本来是一个数学游戏问题，但聪颖过人的欧拉并不限于处理这种特殊情形，而是以此为基础，研究了当时超出欧氏几何范围的几何问题。这使得欧拉成为图论和拓扑学的创始人，这是始料不及的，因为当时人们并没有体会到七座桥的图论和拓扑学意义。

二 四色问题

许多学者在对图论或拓扑学的研究中，比七座桥更感兴趣的是著名的“四色问题”。

什么是四色问题？它是怎么产生的？以地球仪为例，它的表面被划分为国家、海洋等区域。为了明白显示不同的区域，通常在每一区域上着一种颜色，且要求任何两个相邻的区域需着不同的颜色，称之为可接受的着色。现在的问题是：至少需要几种颜色，才能作出可接受的着色？由此引出四色问题，即对于平面或球面的任何地图的着色中，不管有多少国家，也不管它们怎样设置，至多用四种不同的颜色对各个国家着色，就可以使相邻的国家有不同的颜色。所谓地图上的国家是指单连通区域，有公共边界（而不是只有一个公共点）的国家称为相邻的国家。这实际上是一个图论问题，因为如果用点代表国家，当且仅当两个国家相邻时，在代表这两个国家的点之间连一条线，那么地图四色猜想就变成为图的四色猜想：对于任何一个平面图，至多只用四种颜色对各点着色，就可以使相邻的点具有不同的颜色。

用四种颜色对地图着色虽然只是猜想，但到目前为止，尚未发现需要四种以上颜色才能着色的地图。但要使猜想成为定理，还需要严格的数学证明。很多数学家低估了证明的难度，试证都没有成功。例如素以谦逊著称的德国数学家闵可夫斯基在大学给学生上拓扑课时，拿起粉笔竟要当堂给学生推导出来，结果没有成功。下一节课又去试证，仍没成功，过了几天亦无进展。有一天，他刚跨进教室，适逢天上雷声大作，震耳欲聋，他马上对学生说：“上天在责我自大，我也没法解决四色问题。”这样，四色问题就成了国际上最著名的数学难题之一。直到一百多年前，一位名叫肯普的数学家发表了一篇文章，证明四色的答案是正确的。但遗憾的是，到 1890 年，数学家海伍德发现肯普的推理中有一个错误，后来许多人设法弥补那个错误，但没有人能成功。虽如此，那篇文章的价值仍旧存在，肯普的思考路线带给人们一些启示。例如海伍德利用肯普的思考路线证明了“五色定理”，即证明了对所有的地图着色，五种颜色是足够的。

四色问题不涉及区域的形态，或弧和线段的长度，所以也是属于拓扑学的问题。近数十年来，在拓扑学的蓬勃发展中，不知道有多少拓扑学工作者在探索四色猜想的证明，但直到 1976 年才由美国两位数学家阿佩尔、哈肯和一位计算机专家莫尔合作，利用电子计算机算了 1 200 个小时，作了近 100 亿个逻辑判断，从而完成了四色猜想的证明。但这并不意味着研究工作的结束，这是因为

还有人准备用不同的计算机算法去证明四色猜想，也还因为人们可能会越来越多地依靠计算机来解决数学中的大量逻辑判定和数值计算问题。在这方面我国著名数学家吴文俊的工作一直在世界上处于领先地位，对用机器证明数学定理做出了卓越的贡献。另外，还有许多图论或拓扑学工作者仍在寻求不借助于计算机的证明方法，这是因为四色猜想的证明，其意义远远超过了这个问题本身。为了解决这个问题而提出的各种方法，揭示了图论和拓扑学的许多基本性质，甚至有人把四色问题喻为图论的“催化剂”。□

11 弈棋、计算机和对策论*

弈棋、打牌等都是人们在繁忙的工作和学习之暇所喜爱的文娱活动，也是一种高级的思维活动。这些活动，总是可以把参加者分为对立的并拥有决策权的两方，叫做局中人。局中人既可以理解为个人，也可以理解为集体（如代表两个国家的球队）。如果局中人想在活动中取胜，就要充分运用自己的智谋，扬长避短，选择最优策略。用数学的观点和方法来研究冲突局势下有利害关系的双方是否存在一方制胜另一方的最优策略，以及如何找到这种最优策略的学问，在数学上叫做对策论（Game Theory）。由于这种问题最初是从赌博和弈棋中提出来的，因此对策论也叫做博弈论。它是运筹学的一个重要分支。

对策是竞争者运用策略对抗的过程；策略是指由一系列行动所构成的具体措施。如只有两个局中人，则称为两人对策；如有 $n(n>2)$ 个局中人，则称为 n 人（或多人）对策。对策论的基本思想是立足于最坏的情况，争取最好的结果，即“我”就是赢不了“你”，也力争不输或少输于“你”。

计算机的诞生，是科学发展史上的一个里程碑。计算机不但能够进行高速度运算，而且能够进行思维活动，它开创了人类部分脑力劳动的自动化，丰富了人类的精神财富，提高了人类认识世界的能力。如 50 年代初期，美国 IBM 公司的工程课题研究组成员塞缪尔（Samuel）创造了具有自改善自适应能力的跳棋机。他利用对策理论和启发式搜索技术在 IBM704 计算机上编制了一套跳棋程序，

---

* 本文原载于《电脑爱好者》1993 年第 9 期。

这种程序不是让机器做穷尽一切可能的探索，而是像一个优秀的棋手一样，能向前看五六步，并应用棋手们在实践中摸索到的一些规则，对每一步棋进行评价。1959 年，跳棋机战胜了塞缪尔本人；1962 年，又一举击败了美国一个州的国际跳棋冠军尼亚莱（Nealey）。近年来，随着计算机科学和技术的发展，计算机下棋的本领又有了显著的提高。研制博弈计算机，不仅仅是计算机科学家和工程师们感兴趣的研究课题，它也引起了军事家、数学家和哲学家们的巨大关注。

弈棋在数学上称为博弈，是研究对策的，把双方的活动用数学语言描述，寻找最优的策略。由于走一步棋有多种多样的选择，所以需要确定一种选择标准，通常是参照名家棋法和棋谱作为走下一步的标准。

如果要让计算机会下棋，就需要先将下棋的博弈规则用算法语言编成程序，存储在计算机里。计算机要成为一个下棋能手，也并非易事。一方面要求计算机具有高速运算能力，另一方面还要求计算机有足够大的存储空间，能存储大量棋谱。除此之外，还要求计算机具有“学习”能力，能在各种场合下对出现的情况进行比较、分析和判断，这就是所谓的“智能模拟”。智能是人脑的一种机能，它包括学习、记忆、联想和逻辑思维过程等。运用电脑学习机便可模拟人脑的学习、记忆和思维过程。电脑学习机的构造与人脑没有什么共同之处，但是它可以完成人脑才能够完成的学习、记忆、数学运算、逻辑推理等功能，而且速度之快，准确度之高，常常不是人脑所能达到的。这种方法称为智能模拟方法，也就是人工智能。它的主要内容就是探索和模拟人的感觉以及思维过程，它的产物之一就是智能机器人。现在对智能机器人的要求就是能识别图像，能理解人的语言，能学习，能对事物进行探索、联想，能证明定理，能求解问题等等。目前，国外已研制成一些智能机器人的样机。如美国斯坦福大学研究所研制的能看、能记忆、能走路、看到障碍物会自动避开、走过一次能记住所走路线的智能机器人。

对策论的实际应用在 20 世纪 50 年代和 60 年代有了长足的进展，各种对策论的文章如雨后春笋般地涌现出来。动态规划奠基人、美国数学家贝尔曼（Bellman）在自动控制系统中应用了博弈论，使动态规划成为研究最优控制的一种重要方法。他的名著《博弈、动态规划和计算机》获得了广泛的好评，推动了各种对策的深入研究。一开始，军事家运用对策论思想来研究水雷和军舰、歼击机和轰炸机、导弹和反导弹之间的作战和追踪问题，后来，航天技术中出现了机动追击的对策问题，原来的对策论就难以适应。于是，1965 年，美国著名学者、数学家依萨克斯（Isaacs）提出了在追踪问题中逃、躲双方都能自由决策行动的微分对策理论，从而使对策论的军事应用进入了一个新的发展历程。

电子计算机是实现国防现代化的一种重要工具，它已成为反导弹武器系统中的枢纽。电子计算机主要用来控制武器对目标进行准确射击，比如，地对空导弹可用来对付来袭的敌机，由于敌机和导弹飞行的速度都很快，需要用电子计算机来操纵导弹，使它一直飞到离敌机很近处爆炸。操纵的办法可以用两组雷达，一组跟踪敌机，一组跟踪发射出去的导弹，两组雷达分别不断地确定敌机和导弹的位置，把测得的数据发送到计算机上；计算机接收到雷达的数据之后，就自动地进行计算，算出下一时刻飞机的位置，修正导弹飞行路线的数据。这些数据由指挥站用无线电波形式发送给导弹，导弹通过自带的自动控制系统和自动驾驶仪，把这些信号转换成操纵指令，准确命中敌机。进一步还可以把雷达和电子计算机都装在导弹里面，使导弹能够自己探测敌机的位置，计算敌机下一时刻的位置和自己应当怎样纠正飞行路线，并且自动地纠正自己的路线。

在这种导弹里的电子计算机还有一个很重要的性能，就是能抵抗干扰。因为敌机会放出信号来干扰导弹，使导弹不能命中，所以导弹要能抵抗干扰。在这种情况下，敌机又会发出信号来干扰导弹的抗干扰，因此，导弹又要抵抗这种对于导弹抗干扰的干扰。这种情形和下棋类似：你走了一步，我就要走一步来抵抗你的一步。因此，可以根据博弈原理来设计控制导弹的电子计算机和自动控制系统，使它能够抗干扰。

对策论的应用还在不断地发展。□

12 对策论*

在日常生活中人们的下象棋、玩扑克，在空战中歼击机追踪轰炸机，在生物界中猫追老鼠、猎犬追野兔等，这都是所谓的“竞技”行为。竞争或斗争的一方都希望最终能获得胜利，因此必须考虑对方可能怎样决策，力争利用对方弱点，发挥自己的优点，选出一个最优策略，以战胜对方。对策论就是利用数学方法研究竞争或斗争中是否存在一方制胜另一方的最优策略及如何找到最优策略的一门应用数学。

对策论的朴素思想在我国可谓源远流长。例如，在公元前4世纪的战国时期，就有齐王与大将田忌赛马的故事。田忌与齐王的马都分上、中、下三个等级，但同级的马，齐王的要比田忌的稍强。两人约定：(1) 双方各从自己三个等级的马中选出一匹参加比赛；(2) 每匹马都参加且只能参加一次比赛，一共比赛3次；(3) 每次比赛后败者要付给胜者千金。看来田忌要输三千金已成定局。但田忌的谋士孙膑向田忌献策：(1) 每次比赛先让齐王说出他要出哪匹马；(2) 用你的下马对齐王的上马，上马对齐王的中马，中马对齐王的下马。比赛结果，田忌两胜一负，反而赢了千金。这是一个典型对策问题。但对策论真正形成为一门独立学科，还是在1944年冯·诺伊曼和摩根斯坦(Morgenstein)合著的《对策论与经济行为》一书出版以后的事。这本书也是对策论的奠基之作。1946年以后，由于

*本文为作者根据《科技发展与改革》1989年第11期《追逃的智谋——微分对策》和《自动化》1988年第6期《对策论简述》两文整理。

计算机的出现和利用，大大简化了对策论中的复杂计算，才使它不是纸上谈兵。20 世纪 60 年代后，对策论与最优控制相互结合并使其得到了飞速的发展。1965 年，美国学者依萨克斯出版了《微分对策》一书，正式创立了微分对策。同年，美国自动控制专家扎德发表了题为《模糊集合》的论文，从而确立了模糊数学。对策论在模糊数学中也找到了它的应用，出现了“模糊对策”，但目前比较成熟且最基本的还是矩阵对策。

一 矩阵对策

对策论一般具有三个最基本的要素，即局中人、策略集和赢得或支付函数。直接参与对策的一方叫局中人。只有两方的对策叫两人对策，多于两个局中人的对策叫多人对策。一个局中人从开始到终了的完整行动方案叫做局中人策略。策略的全体构成策略集合。如果局中人只有有限个策略就叫有限对策，否则就称为无限对策。一局对策结束之后，胜的一方赢得或败的一方损失（或支付）统称为得失。在两人对策中胜的一方之所得，就是败的一方之所失，即两个局中人得失之和为零，称这样的对策为两人零和对策。由各个局中人的策略集合中各取一个策略所组成的策略组叫局势。一局之得失是局势的函数，叫赢得（或支付）函数。齐王赢得千金，田忌就失掉千金，所以齐王与田忌的赛马是两人有限零和对策。以策略 α_1（上中下）表示齐王先出上马，次出中马，后出下马；以策略 β_1（上中下）表示田忌先出上马，后出中马，最后出下马……以 1 和 3 表示齐王赢得的千金数，同时也是田忌应付的千金数。负值是齐王应付的千金数，也是田忌应得的千金数。齐王赢得如表 1 所示。

表 1

	β_1（上中下）	β_2（上下中）	β_3（中上下）	β_4（中下上）	β_5（下中上）	β_6（下上中）
α_1（上中下）	3	1	1	1	1	−1
α_2（上下中）	1	3	1	1	−1	1
α_3（中上下）	1	−1	3	1	1	1
α_4（中下上）	−1	1	1	3	1	1
α_5（下中上）	1	1	−1	1	3	1
α_6（下上中）	1	1	1	−1	1	3

实际上，这张表是一个 6×6 矩阵。齐王的策略集是行，田忌的策略集是列，称此矩阵为齐王的赢得矩阵。因为齐王之所得就是田忌之所失，所以田忌的赢得矩阵就是齐王赢得矩阵中各元素变号而成的矩阵。于是任何一个两人零和有限对策给定之后，其表示对策结果的数值矩阵就唯一地确定。反之，任何一个数值矩阵都可看成是一个两人零和有限对策，因此两人零和有限对策也称为矩阵对策。

在矩阵对策中，局中人怎样选择最优策略呢？例如，设有一个两人有限零和对策，局中人甲、乙的策略集及甲的赢得如表 2，求两人最优策略和赢得。

表 2

	β_1	β_2^*	β_3	各行的最小数
α_1	−6	1	−8	−8
α_2^*	5	3	4	3*
α_3	8	−1	−9	−9
α_4	−1	0	5	−1
各行的最大数	8	3*	5	

在研究两人对策时基本思路是立足于最坏的情况，争取最好的结果。为了清楚，我们先写出矩阵：

$$\boldsymbol{A}=\begin{pmatrix}-6 & 1 & -8\\ 5 & 3 & 4\\ 8 & -1 & -9\\ -1 & 0 & 5\end{pmatrix}。$$

对局中人甲来说，所有最坏情况为各策略所对应行的最小数：−8，3，−9，−1。再从这些最坏的情况中寻求最好的结果，即 3。对局中人乙来说，所有最坏的结果即 $\boldsymbol{A}$ 中的每一列的最大元素：8，3，5。在这些最坏的结果中，最好的情况也是 3。以 a_{ij} 代表 A 的任一元素，把以上情况综合起来写则有

$$\max_i\{\min_j\{a_{ij}\}\}=\min_j\{\max_i\{a_{ij}\}\}=3。\tag{1}$$

它对应于甲的策略是α_2^*，对应于乙的策略是β_2^*。只要甲选取α_2^*参加对策，就保证赢得不少于 3；同样，只有乙选取β_2^*来参加对策，就保证支付不会大于 3。因此，α_2^*，β_2^* 分别叫做甲和乙的最优策略。

一般地，设矩阵对策 $G=\{S_1,S_2,\boldsymbol{A}\}$，其中 $S_1=\{\alpha_1,\cdots,\alpha_m\}$ 和 $S_2=\{\beta_1,\cdots,\beta_n\}$ 分别为甲、乙的策略集，$\boldsymbol{A}=(a_{ij})_{m\times n}$ 为甲的赢得矩阵。如果等式(1)成立，记

其值为 V_G，则称对策有解，并把 V_G 叫做对策的值。策略 α_i^* 和 β_j^* 称为它们各自的最优策略，(α_i^*, β_j^*) 称为这个对策的解。它既是甲的最大值，又是乙的最小值。在几何上，(α_i^*, β_j^*) 的位置似马鞍的最低点，因此称为鞍点。

二　混合策略

如果矩阵对策有鞍点存在，局中人应选择鞍点所在的行和列所对应的策略。但并不是在一切矩阵对策中都有一个鞍点，如局中人甲的赢得如表 3。

表 3

	β_1	β_2	各行的最小数
α_1	1	3	1
α_2	4	2	2^*
各列的最大数	4	3^*	

这样，新列中的最大数 2^* 不等于新行中最小数 3^*，于是鞍点不存在。局中人如何来选择策略参加对策呢？在一般的对策中，各个局中人出什么策略必须防止被对方猜到，因此每个局中人就得要以某固定的概率随机地选取策略，这样就产生了混合策略，它是前述策略的扩充。为避免混淆，称前述策略为纯策略。现假定甲以概率 x 选取策略 α_1，以概率 $(1-x)$ 选取策略 α_2。又假定乙以概率 y 选取策略 β_1，以概率 $(1-y)$ 选取策略 β_2。于是，对甲来说他的期望赢得为

$$E(x,y)=1xy+3x(1-y)+4(1-x)y+2(1-x)(1-y)$$
$$=-4(x-\frac{1}{2})(y-\frac{1}{4})+\frac{5}{2}。$$

由此式可以看出，当 $x=\frac{1}{2}$ 时，$E(x,y)=\frac{5}{2}$。就是说，当甲以概率 $\frac{1}{2}$ 选取策略 α_1 时，他的期望赢得至少是 $\frac{5}{2}$。同样，乙只有取 $y=\frac{1}{4}$ 时才能保证他的支付不会多于 $\frac{5}{2}$。于是甲以概率 $x_1=\frac{1}{2}, x_2=\frac{1}{2}$ 选取 α_1, α_2，乙以概率 $y_1=\frac{1}{4}, y_2=\frac{3}{4}$ 选取 β_1, β_2 时，双方都能得到满意的结果。

我们把甲、乙分别以概率 $x_1=x_2=\frac{1}{2}$ 和 $y_1=\frac{1}{4}, y_2=\frac{3}{4}$ 选取 α_1, α_2 及 β_1, β_2 写成向量的形式

$$X=(x_1,x_2)=(\frac{1}{2},\frac{1}{2}),x_1+x_2=1;$$

$$Y=(y_1,y_2)=(\frac{1}{4},\frac{3}{4}),y_1+y_2=1。$$

这样一组按顺序排列的概率也算是一个策略，它是前述策略的扩充，叫混合策略。

一般地，设局中人甲有 m 个纯策略，我们把甲的纯策略集所对应的概率写成

$$S_甲=\{X=(x_1,\cdots,x_m),x_i\geqslant 0,i=1,2,\cdots,m,\ \sum_{i=1}^{m}x_i=1\},$$

$S_甲$ 叫做甲的混合策略集。

同样地，设局中人乙有 n 个纯策略，把乙的纯策略集所对应的概率写成

$$S_乙=\{Y=(y_1,\cdots,y_n),y_j\geqslant 0,j=1,2,\cdots,n,\ \sum_{j=1}^{n}y_j=1\},$$

$S_乙$ 叫做乙的混合策略集。

当甲选取策略 $X=(x_1,\cdots,x_m)$，乙取策略 $Y=(y_1,\cdots,y_n)$ 时，甲的赢得定义为赢得的期望值，即

$$E(X,Y)=\sum_{j=1}^{n}\sum_{i=1}^{m}a_{ij}x_iy_j,$$

其中，a_{ij} 为原赢得矩阵的元素。乙的赢得定义为 $-E(X,Y)$。

在对策论上，对任意赢得矩阵 $A=(a_{ij})_{m\times n}$，都有

$$\max_{X\in S_甲}\min_{Y\in S_乙}E(X,Y)=\min_{Y\in S_甲}\max_{X\in S_乙}E(X,Y), \tag{2}$$

这就是著名的冯·诺伊曼和摩根斯坦的极大极小定理。

以 V 代表(2)式的数值，叫做矩阵对策 A 的值。由等式(2)可推出必有混合策略 X^*,Y^*，使

$$\max_{Y\in S_乙}E(X^*,Y)=\max_{X\in S_甲}E(X,Y^*) \tag{3}$$

成立。当甲取策略 X^* 时，就保证自己的赢得不会少于 V；当乙取策略 Y^* 时，就保证自己的支付至多是 V。使(3)成立的 X^* 和 Y^* 的全体分别叫做甲和乙的最优混合策略集。混合局势(X^*,Y^*)称为对策的混合扩充解。应用矩阵运算和电子计算机就可以把这些最优混合策略全部求出来。不过，这里的最优策略乃是取策略的最合适的概率，因此它与前述最优策略是不一样的。

三　微分对策

微分对策起源于军事斗争中的“追踪”与“逃避”问题，局中人就是追者与逃者。例如，在反坦克作战中，坦克为一方，导弹为另一方；在飞机空战中，歼击机为一方，轰炸机为另一方，等等。由于追逃的情况要用微分方程组来描述，因此在数学上把研究追逃问题的理论称微分对策。

从历史角度看，微分对策理论有两方面的根源：一是对策理论；二是最优控制理论。微分对策与对策论有密切关系，它是对策论的进一步发展。以前的对策论是静态情况下的对策，所以叫静态对策；微分对策是静态对策在动态情况下的扩充，所以是动态对策。由静态对策到动态对策，这在对策理论的发展上是一重大突破，在这方面，美国学者依萨克斯做出了卓越的贡献。20 世纪 40 年代，美国兰德公司的依萨克斯等人出于军事科学的需要，将对策论与军事运筹学结合在一起进行研究，并于 1954 和 1955 年发表了四篇论著。1965 年，他将上述论述整理成专著《微分对策》出版，提出了追踪问题中追逃双方都能决策行动的微分对策理论。1971 年，美国学者弗里德曼在依萨克斯研究的基础上又进一步确立了微分对策的数学理论，并定义了微分对策的一些基本概念，如微分对策的值函数、策略和鞍点等，并发展了基本的存在定理，使微分对策逐渐完善和系统化。

另一方面，微分对策与最优控制也密切相关。与微分对策产生的同时，由于自动控制技术的广泛应用，从而促使了自动控制理论的迅猛发展，这可以看做是一方的控制。人们经过深入研究发现，微分对策实质上是一个双边的最优控制问题，所不同的是最优控制要么求最小值，要么求最大值，而两人微分对策则是寻求最大值中的最小值及最小值中的最大值问题。目前，人们对微分对策的研究方兴未艾，新的微分对策模型不断出现，如两人零和微分对策、两人非零和微分对策、多人微分对策、偏微分对策及随机微分对策等。但总起来说，可分定量与定性两类微分对策，现作简单介绍。

定量微分对策。所谓定量就是使表达某一性能指标的量（如距离、时间等）有极值。例如，设一只猎犬 A 追捕一只野兔 B，在时刻 t，A，B 的位置坐标分别为 (x_1,y_1) 和 (x_2,y_2)，其速度大小分别记为 u 及 v，运动方向与 x 轴的夹角分别为 α 和 β。猎犬为追捕野兔总是朝野兔奔跑，而野兔为不让猎犬追上，则想方设法拉大它们之间的距离，时快时慢，有时还来个大转弯。猎犬也采取相应的对策，随时改变着自己的行动方案。位置坐标 x，y 是描述两个局中人在每一时刻

t 相对位置的变量，叫状态变量。状态变量是时间 t 的函数，由此我们可以得到描述 A,B 运动情况的微分方程组

$$\begin{cases} x_1=|u|\cos\alpha, \\ y_1=|u|\sin\alpha; \end{cases} \quad \begin{cases} x_2=|v|\cos\beta, \\ y_2=|v|\sin\beta。 \end{cases}$$

在追逃过程中，双方的控制手段实际上不能不受到一定的限制。如速度都不可能很大、不可能转很小的弯等。在数学上这些限制可用一组不等式来表示，叫做控制约束条件。如

$$0\leqslant|u|\leqslant|u|_{\max},0\leqslant|v|\leqslant|v|_{\max}。$$

式中 $|u|_{\max}$，$|v|_{\max}$ 都是给定的正数。当 A,B 之间的距离足够接近，用数学式表示就是当

$$(x_1-x_2)^2+(y_1-y_2)^2=J^2$$

时，就认为实现了捕获。式中 $J(J>0)$ 称为捕获距离，又称为赢得或支付。这是一个零和微分对策的问题。对于 A 选 u^*，使捕获距离 J 最小；对于 B 选 v^*，使 J 最大。问题归纳为求最优策略 (u^*,v^*)，并使其满足

$$J(u^*,v)\leqslant J(u^*,v^*)\leqslant J(u,v^*)。$$

(u^*,v^*) 称为鞍点。(u^*,v^*) 必满足双方极值原理，它是极大值原理的推广。

如果以 t_0 表示初始时刻，t_f 表示捕获时刻，则 t_f-t_0 就是猎犬追捕到野兔所需的时间。猎犬要求这段时间尽可能短，而野兔则想方设法延长时间，如果这段时间为无限长，那就表示猎犬永远追不上野兔了。

定性微分对策。定性微分对策是研究在什么条件下对策结束的一种对策。定性微分对策也是研究追逃问题的，但追逃是在状态空间内进行。状态空间可能存在着两个区域：一个叫“捕区”，一个叫“逃区”。当追者在捕区时，不论它采取什么样的策略，追者总可以采取适当的策略使逃者被捕；当逃者在逃区时，不论它采取什么样的策略，它就可以摆脱被捕。逃区与捕区的界面称为壁垒或界栅，如果这个壁垒被找到，问题就得到了明确的答案。如果追逃双方势均力敌，则双方到不了上述的捕区与逃区，这时对策就在界面上进行。设双方各自选其最优策略，这时界面是与此最优策略相对应的最优轨线。

微分对策的方法不仅已应用在军事、公安、自动控制、航天、环境保护及海洋捕捞等方面，而且还进入到了经济、政治领域。但微分对策毕竟还是一门年轻科学，它的理论和应用不论在广度还是在深度方面，都有许多问题值得人们去进一步探讨。例如，模糊微分对策就是一个新的领域。因为人类社会存在着许多模糊策略，可以用模糊数学探讨当约束条件或策略在模糊情况下的微分对策，这有待人们进行深入的研究。□

现代控制理论：研究系统的调节与控制的一般规律

1 控制与信息*

1948年，世界数学巨匠维纳（Norbert Wiener）发表奠基性著作《控制论》，这本书的副标题明确指出控制论的研究对象是“动物和机器中控制和通讯的科学”。著名控制论专家、英国生理学家艾什比（W. R. Ashby）在其《控制论导引》一书中认为控制论也确实可以定义为：它是研究这样一类系统的科学，在系统中能量是无关紧要的，而信息及控制却非常重要。同时，艾什比强调，控制论是研究复杂系统的，现在大规模的复杂生产系统、自动飞行系统、宇航系统、武器系统、科学实验系统等，都是依靠信息进行控制的系统，这种系统在生物界中也屡见不鲜。

和所有的新兴学科一样，控制论的诞生与发展并非偶然，而是有着悠久的思想渊源与深刻的历史背景。远的不谈，在第二次世界大战期间，由于改进防空火力系统的需要，维纳在先后两次从事研究一种用来代替人类特殊功能的机电系统的工作中，用他研制的高炮自动瞄准仪，成功地代替了高炮操作者。维纳从中发现了自动控制装置在行为上同人和动物这样的生命有机体的相似性。他指出，“一切有目的的行为都可以看做是需要负反馈的行为”。他通过“行为”把“反馈”和“目的”联系起来。后来，维纳在《控制论》一书中明确提出控制论的两个基本概念，即信息概念和反馈概念，揭示了机器、生物和人所遵从的共同基本规律——信息与控制规律，为机器模拟人和动物的行为或功能提供了理论根据。

---

* 本文原载于《自动化博览》1995年第4期。

控制是控制论中最重要的概念,它是和目的性直接相关的。没有目的,就谈不上控制;同样,没有信息,也就没有控制。人的有目的性的行为又总是同外界环境进行联系的,这种联系是一种信息联系,即依靠信息的输入和输出,实行自身内部的通讯和保持有机体与外界的平衡。人与外界联系并达到一定目的的手段是反馈,依靠信息反馈对外界对象进行控制。所谓信息反馈,是指输出的一部分信息再输送回来,以便同原来规定的标准值或目标值进行对比,及时发现偏差,加以纠正。比如,一个正常的人用手拿东西,先用眼睛盯着要拿的东西,并不断目测手与东西之间的距离,随时将偏差信号(即信息)反馈给大脑,大脑不断指挥手向既定目标运动,直至拿到东西为止。这种反馈称为"负反馈",即从输出端反馈回来的实际值与期望值进行比较,使差值不断减小或趋近于零。反馈是维纳控制理论的核心概念,是任何一个系统的基本要求,是动物和机器控制的共同特性。现代的自动驾驶系统就是利用反馈信息进行控制的。为了按既定航向自动飞行,首先将期望的航向与实际的航向进行比较,然后根据误差信号操纵飞机的控制部件,以便使航向误差减少到零。

维纳等人关于控制论的一个基本理论观点,就是控制系统也是一种信息系统,因此,必须用信息的观点来研究控制系统。正是由于控制论研究的理论需要,维纳也研究了信息论问题。在维纳创立控制论的同年,美国贝尔电话公司的数学家仙农发表了题为《通信的数学理论》的权威性论文,这种时间上的巧合,说明了控制论和信息论这两个新兴学科之间的内在联系。仙农把通信系统中的消息称为信息,并试图给出这种信息的度量,这种从通信系统中提出的信息概念,被称为狭义信息概念。但究竟什么是信息,至今人们对它还没有一个公认的定义。从来源的角度看,信息就是用以表征客体变化或客体间相互差异或关系的;从认识的角度讲,信息是指主体对客体的不定性的认识程度,即获得知识的程度。仙农把信息理解为一种概率增加。在仙农看来,信息乃是"两次不定性之差"。信息量也就是不定性减少或消除的数量。仙农的这种思想和统计热力学中的熵有着密切的联系。"熵"是一种不定性的数量度量,它表征了一个物理系统的无组织程度。信息是一种被消除了的不定性,因此可以看做是"负熵",并用概率统计的方法,给出了平均信息量的数学表达式

$$H(X)=-K\sum_{i=1}^{n}P(x_i)\log_a P(x_i)。$$

其中,$H(X)$表示某事物(或事件集)X的信息量,K为系数,$P(x_i)$为X事物中某事件x_i可能出现的概率,$i=1,2,\cdots,n$,n为X事物可能出现的事件总数。当

对数底数 a 取 2，且 $n=2, P(x_1)=P(x_2)$ 时，令

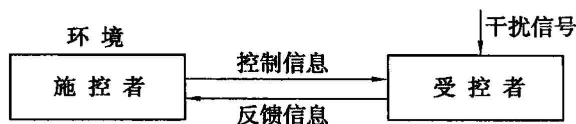
$$H(X)=-K\sum_{i=1}^{2}P(x_i)\log_2 P(x_i)=1。$$

以此作为信息量单位，称为 1 比特(bit)。1 比特信息量就是含有两个独立等概率可能的事件所具有的不确定性被全部消除所需要的信息。

仙农的信息论被称为狭义信息论，是一门应用数理统计方法研究信息的处理和信息传递的科学，其根本目的是提高信息传递系统的有效性与可靠性，基本内容是信源、信宿、信道以及编码等问题。

除了仙农被公认为是信息论的创始人之外，控制论创始人维纳也对信息论做出了不可忽视的贡献。仙农是维纳的学生，在创立信息论的过程中，仙农从维纳那里得到了不少的启示，维纳则是从电滤波器中的噪声与消息问题出发来研究信息的统计理论的。维纳还建立了著名的维纳滤波理论和信号预测理论，阐明了信息定量化的方法和度量公式；维纳又从控制论的角度出发，提出了含义更为广泛的信息概念。在《控制论》一书中，他认为“信息是我们适应外部世界，并且使这种适应为外部世界所感受的过程中，同外部世界进行交换的内容的总名称”。从这个定义可以看出，信息是指人、动物或机器等控制系统同外界联系的一种形式。

信息概念的形成，使我们能够将任意一个控制系统都看做是信息系统，这种系统可用以下框图表示：



我们可以把施控者发出控制指令的信号看做信源发出的消息，而受控者发出的反馈信号看做经译码后所收到的消息，要使得所收到的消息与原来所发出的消息一样，就要对失真加以控制，这与一个系统因扰动而偏离目标必须加以控制是相似的。控制系统要有信息处理才能达到目的，而通讯系统要有效而可靠，也离不开控制，所以这二者是很难加以区分的。从这样一个基本观点出发，可以认为控制论所说的反馈，主要是信息反馈。

总之，控制论是一门研究利用信息来进行调节和控制的科学。信息论着眼于对信息的认识(描述和度量)，控制论着眼于对信息的利用(处理和控制)。如同人类认识世界的目的是为了改造世界一样，人们认识信息、处理信息，也往往是为了对系统进行控制。随着信息概念的拓广，控制论的研究范围也将更为广

泛，一旦广义信息的定量化方面能够取得实质性的进展，将会产生新的控制系统。所谓广义信息，是相对于仙农的狭义信息而言，这些信息包含人类感官所能感知的一切有意义的东西，它是事物存在的方式（或运动状态）以及这种方式（运动状态）的直接或间接的表述。目前，关于广义信息的研究还仅处于初始阶段，它还在探索和发展之中。□

2 变分法、最小值原理、动态规划和最优控制*

变分法、最小值原理和动态规划三者都是研究优化问题的，而且也是求解最优控制的有力工具。

从数学上讲，最优就是寻求函数的极值（极小或极大值）问题。电子计算机的出现和发展，把最优化推进到了一个崭新阶段。

17 世纪，微积分的创立，从根本上推动了极值问题的研究。设多元函数 $y=f(x_1,x_2,\cdots,x_n)$ 在某个开区间连续可微，求其极值时，先是求 y 的全微分，得到

$$dy=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\cdots+\frac{\partial f}{\partial x_n}dx_n,$$

然后再令 $dy=0$，即分别令 $\frac{\partial f}{\partial x_1}=0,\cdots,\frac{\partial f}{\partial x_n}=0$，联立求解，即得该函数极值的一组必要条件（非充分条件）。至于究竟是极大还是极小，则需要求二次微分 d^2y。于是，函数求极值问题就主要归结为求解方程组问题。如还带有约束条件，则可通过拉格朗日乘子法，变有约束求极值问题为无约束求极值问题。

所谓泛函，可以看做是普通函数的推广。设一个变量 J，如果对于某一类函数的每一个函数 $x(t)$，都有一个确定的 J 的值与之相对应，则称 J 为依赖于函数 $x(t)$ 的变分，记为 $J=[x(t)]$。这里的自变量仍是一个函数，故泛函可以简单地理解为“函数的函数”。研究泛函极值的方法称为变分法或变分学。如同函数 $y=f(x)$ 的微分是函数 y

---

* 本文原载于《自动化博览》1997 年第 6 期。

的增量$\triangle y=y'(x)\triangle x+r(x,\triangle x)$，当$\triangle x\to 0$时，第一项叫$\triangle y$的线性主部，通常叫函数$y$的微分$dy=f(x)\triangle x$；在泛函$J=[x(t)]$中，泛函$J$的增量$\triangle J=J(x+\delta x)-J(x)=L[x,\delta x]+r[x,\delta x]$，其中第一项是泛函数增量的线性主部，第二项是关于$\delta x$的高阶无穷小(当$\delta x\to 0$时)。通常，把第一项叫做泛函$J$的变分，记为$\delta J=L[x,\delta x]$，所以泛函的变分也可称为泛函的微分。类似于求多元函数$y=F(x_1,\cdots,x_n)$的极值，在变分学中求多元泛函$J=[x_1(t),\cdots,x_n(t)]$的极值时，首先是求J的变分，即$\delta J=\frac{\partial F}{\partial x_1}\delta x_1+\frac{\partial F}{\partial x_1}\delta x_2+\cdots+\frac{\partial F}{\partial x_n}\delta x_n$，然后令$\delta J=0$，即得泛函$J$的极值函数。为什么是极值函数而不是极值呢？这是因为$x_1=x_1(t),x_2=x_2(t),\cdots$都是关于$t$的函数。因此，求泛函的极值将面临求解一组微分方程的问题，而微分方程的解一般是一个函数，而不是数值。同时，求解微分方程还涉及到确定积分常数的问题，也即涉及边界条件(初始条件)的问题。不难看出，求泛函极值比求函数极值要复杂得多。同样，一次变分$\delta J=0$，只是求泛函极值的必要条件(非充分条件)，要想进一步判断极值函数是极大还是极小，则需求二次变分$\delta^2 J$，且当$\delta^2 J>0$时，$J=\min$(极小)，而$\delta^2 J<0$时，$J=\max$(极大)。

泛函经常是以定积分的形式出现的。在最优控制中，泛函可以用来泛指控制系统期望达到的目标、指标或准则，因此常称为“目标函数”、“性能指标”等。

为了区别于近代发展起来的求极值方法(如最小值原理等)，通常又称变分法为古典变分法。古典变分法只能解决一些比较简单的最优控制问题，试图用这种方法去解比较复杂的问题就会遇到很大的困难。因为最优控制所要解决的问题是，按照控制对象的动态特性选择一个容许控制，使得被控制对象按照技术要求运转，同时使某一性能指标达到最优值。从数学方面看，就是求解一类带有约束条件的泛函极值问题。因此，这是一个变分学问题。然而变分理论只是解决容许控制属于开集的一类最优控制问题，而在工程实践中更多的却是容许控制属于闭集的另一类最优控制问题，对这类问题，古典变分法是无能为力的，所以变分法虽然是在最优控制理论中占有重要地位，但是用古典变分法来解决最优控制问题是有一定困难的，这就要求人们研究新方法。在研究最优控制理论的解法中，有两种方法最有成效：一种是前苏联数学家庞特里亚金(Лев Семёнович Понтрягин)等人创立的最小值原理，另一种是美国数学家贝尔曼(Bellman)提出的动态规划法。

所谓最小值原理是指当控制作用μ的大小限制在一定范围内时，由最优控制规律所确定的最优轨线在整个范围内必取最小值。

最小值原理也称最大值原理。从数学上来讲，最小值原理与最大值原理没

有什么本质上的区别，因为对任意函数 $H=\min$，也可以描述为 $-H=\max$。

最小值原理是庞特里亚金受力学中哈密顿(Hamilton)原理的启发，在 1956 年到 1958 年逐步创立的。先是推测出最小值原理的结论，随后鲍尔强斯基运用马克申(Mcshane)变分和凸锥分离定理证明了这一原理。最小值原理是在总结并运用古典变分法的基础上发展起来的，它最初应用于时间连续系统，后来又推广到时间离散系统。这种方法的一个显著特点是易于确定最优控制系统的普遍结构形式，因而应用很广，成为求解最优控制问题的一个强有力的工具。

庞特里亚金最小值原理是解决最优控制问题的理论根据，然而最小值原理只给出最优控制的必要条件，而不是充分条件，即并非满足最小值原理的所有控制都是最优控制。但可以证明，对线性系统来说，最小值原理既是泛函 J 取极小值的必要条件，同时也是充分条件。

庞特里亚金等人把最小值原理作为推测提出来之后，不久就给出了严格的论证，其中包含一系列的定理，证明过程十分复杂。稍后，在 1954 年，罗宋诺尔(Rozonoer)利用增量法对最小值原理提供了一个浅显的证明，需要的数学知识较少，“不图严密，但求实用”。因此，一般采用的证法是以增量法为主的。

古典变分法一般是指在没有约束条件或在开集性约束条件下，通过计算变分来计算泛函极值的方法。最小值原理是古典变分法的发展，它能解决古典变分法难以解决的问题，即不仅可用来求解容许控制属于开集的一类控制问题，而且也可用来求解容许控制属于闭集的一类控制问题，也就是把变分法从古典变分阶段推进到非古典变分(即现代变分)阶段。这是在最优控制理论方面的一项重大突破，具有重大的历史意义和现实意义。

动态规划是美国数学家贝尔曼在 20 世纪 50 年代末为了研究多级决策问题而提出来的一种决策过程优化方法，其中心思想是将一个多级决策问题巧妙地转化为多个一级决策的问题，并创立了一套由后向前的“逆向递推”算法。这种方法非常适合于计算机的计算，高速电子计算机的出现，更使动态规划的具体实现有了保证。

贝尔曼研究了这样一种多阶段决策过程：对过程中任意时刻 t，过程在时刻 t 以后的行为，仅依赖于 t 时刻的状态，而与 t 时刻之前过程的行为以及达到这一状态的方式无关，这就是所谓过程的无后效性。针对无后效的多阶段决策过程，贝尔曼提出了一条朴素的原理——最优性原理。它的含义是：一个最优策略具有这样的性质，不论过去的状态和策略如何，如把现在的状态当作后续过程的初始状态，则其后续决策必须构成一个最优策略。换言之，整体的策略最优时，每一级的策略也必须最优。最优的策略得以成立的一个前提条件，是所谓过程的无后效性，其数学描述则是贝尔曼逆推方程。

其实，最优性原理就是“化整为零”的思想体现。它将一个原来需要把全部可能的方案都罗列出来，最后才能作出决断的问题，转化为在每一个关键的地方作出决策的问题，也就是把一个多级决策的问题转化为多个一级决策的问题，从而使一个复杂的问题变得简单化了。

与最小值原理相反，动态规划最初应用于时间离散系统，即多阶段决策问题，随后又推广到了时间连续系统。连续系统的动态规划是在古典哈密顿—雅可比(Hamilton—Jacobi)理论的基础上发展起来的。变分学中古典哈密顿—雅可比理论，研究的是关于寻找满足最优效益的偏微分方程，以及同时存在着的满足最优决策函数的向量偏微分方程的问题(这种方程的求解实际上是困难的，一般都需要求助于计算机)。贝尔曼依据最优性原理，把哈密顿—雅可比理论推进了一步，建立了哈密顿—雅可比—贝尔曼方程。总之，最优性原理是动态规划的基本原理，逆向递推法是动态规划的基本方法。其实，逆向递推法还是奠基于最优性原理之上的。有人形象地把多阶段决策过程比喻为一根链条，一环紧扣一环，为了求得一个最优策略，必须先解开最后一环，逐步反推回去，直到起始状态为止，所得的解就是整个问题的最优解。正因为最优性原理反映了决策过程最优化的本质，故而动态规划得以成功地应用于众多领域，许多问题用动态规划的方法去处理，常比线性规划或非线性规划更为有效。特别是对于离散的问题，由于解析数学无法施展其术，而动态规划的方法就成为非常有用的工具，它不仅可用来求解许多动态最优化问题，而且还可用来求解某些静态最优化问题。

当然，十全十美的方法是少见的，动态规划也有缺点，其对建立数学模型要求较高的技巧，同时也受到维数的限制。维数一高，计算机往往不堪重负，这就是众所周知的“维数灾”。

变分法、最小值原理和动态规划三者研究的是同一类问题，因此必然存在着某种内在联系。

第二次世界大战后，由于航天事业的迫切需要，提出了一类极值控制问题。这类问题的一个共同特点，是它们的控制集或决策集为闭集，故而不能用古典变分法求解。这一困难促使庞特里亚金首先推测出并随后加以严格论证的最小值原理。可以说，最小值原理是以哈密顿方式发展了古典变分法，得到的是常微分方程；而贝尔曼的动态规划可以认为是以哈密顿—雅可比方式发展了古典变分法，得到的是偏微分方程。解常微分方程一般比解偏微分方程容易，但动态规划的应用范围更广泛，且得到的结论是充分条件，所以有可能把最小值原理和古典变分法中的相关条件推演出来。最小值原理的应用范畴比古典变分法广泛些，它既适用于非线性系统，也适用于线性系统；既适用于控制为开集的

情况，也适用于控制为闭集的情况；既适用于边界条件为固定的情况，也适用于边界条件自由的情况。最小值原理的核心是哈密顿函数。哈密顿函数一旦建立起来，其他问题就迎刃而解，所以说最小值原理是古典变分法的推广。当然，变分法也在发展。有人已经发展了古典方法用以解决控制函数受到闭集性约束的问题，但尚不如最小值原理简便。其次，当控制域为开集时，最小值原理与从变分法推出的极值必要条件是等价的，但当控制域为闭集时，最小值原理的优越性就突出了。因为在最优控制与许多应用问题中，控制域往往是闭集，而最小值原理适用于闭集约束，比起古典变分法当然是个重大发展。

动态规划的连续形式与变分法及最小值原理相辅相成，使对控制问题的研究能更深入一步。□

3 最优控制*

第二次世界大战后发展起来的古典控制理论，是以传递函数模型来表示系统的特征，对系统进行分析的。实践证明，对于单输入、单输出的线性定常系统来说，这种方法是很合适的，解决了许多实际工程问题。然而近代航空及空间技术的发展，如导弹的导航、宇宙飞船的对接及现代工业过程等大型复杂系统的出现，不但使控制对象是多输入、多输出的，且参数也是时变的。建立在传递函数基础上的古典控制理论就明显地不能满足需要了。1956 年，前苏联数学家庞特里亚金第一次发表了关于最优控制理论的文章《关于最优过程理论》，揭开了最优控制理论研究的序幕。在此基础上他又加以综合、精炼，于 1961 年出版了《最优过程的数学理论》，成功地建立了最小(大)值原理，解决了以常微分方程为模型的线性系统的最优控制问题，将古典的变分推广到非古典的变分，这是控制理论的重大突破，在一定程度上标志着最优控制理论的诞生。与此同时，50 年代中期美国数学家贝尔曼依据最优性原理发展了变分学中的哈密顿—雅可比的理论，建立了动态规划方法。由于计算机技术的发展，使设计计算和实现控制有了实际可用的工具，为实际应用一些更完善的数学方法提供了工程实现的物质条件。高速度、大容量计算机的应用，一方面使控制理论的工程实现有了可能，另一方面又提出了许多需解决的理论课题。所有这些不仅进一步奠定了最优控制的理论基础，也使这门学科目前得以不断发展。

---

* 本文原载于《自动化博览》1994 年第 1 期。

最优控制理论的内容，概括地说就是在系统的状态方程和约束条件给定的情况下，寻求最优控制的规律（或最优控制策略），使衡量系统的性能指标达到最优，即取得最小值或最大值。所谓最优，一般是指某一性能指标为最优，而不是指任何指标为最优。对于某一性能指标来说，往往要求这一指标的最小值。例如控制时间为最短，燃料消耗为最少等。故最优控制可以更确切地理解为一种极值控制。

为了实现最优控制，首先要建立最优控制的数学模型，但这是最优控制的难题之一。最优控制的一个典型例子就是登月火箭到达月球表面时的软着陆问题。火箭飞行的最后阶段进入了月球引力范围，当火箭垂直自由降落到距离月球表面为 h（约有几百米）的地方（如图），要求火箭速度为零，求最优控制规律，使着陆过程消耗燃料最少。

设火箭质量（含燃料）为 $m(t)$，在火箭速度降为零以前推力为

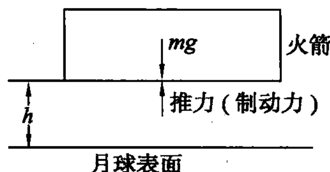
$$F=-K\frac{\mathrm{d}m}{\mathrm{d}t},$$

它起着制动的作用，其中$\frac{\mathrm{d}m}{\mathrm{d}t}$为燃料消耗率，$K$ 为比例常数。火箭从 $t=t_0$ 开始减速，到 $t=t_f$ 速度为零，$X(t_f)=h$ 这一过程火箭的运动方程为

$$m\frac{\mathrm{d}^2X}{\mathrm{d}t^2}=-K\frac{\mathrm{d}m}{\mathrm{d}t}-mg,$$

其中 mg 为月球引力。上面方程称为系统的状态方程。

当 $t=t_0$ 时，$m(t_0)=m_0$ 为初始条件；当 $t=t_f$ 时，$m(t_f)=m_f$ 为终端条件；燃料消耗限制：$-\alpha\leqslant\frac{\mathrm{d}m}{\mathrm{d}t}\leqslant 0$ 为约束条件。



登月火箭到月球表面软着陆示意图

要确定$\frac{\mathrm{d}m}{\mathrm{d}t}$在满足约束条件下，使火箭制动阶段燃料消耗量为最小。这个问题从另一角度看，性能指标燃料消耗量为最小，也即使火箭在着陆时的质量为最大。

从以上例子可看出，最优控制问题具有如下四个“要素”：

（1）受控对象为动态系统。状态方程反映了受控动态系统的运动规律，一般可用动态微分方程或状态方程

$$\dot{x}(t)=f(x(t),u(t),t)$$

来描述，称为受控系统的数学模型。其中 $x(t)$表示 n 维状态向量，$u(t)$表示 m 维控制向量。

（2）初始和终端条件（状态）。动态系统运动过程就是在状态空间中，从一

个状态到另一个状态的转移。如果我们把这种转移看成是 n 维状态空间中点的运动，那么一个动态过程就对应于状态空间中的一条轨线。在初始时刻的初始状态通常是已知的，即 $x(t_0)=x_0$，而所有这些情况都可用一个目标集 S 来加以概括，即 $x(t_f)$ 属于 S。如果终端状态不受约束，则可看做是目标集扩展到整个 n 维空间。

（3）容许控制。控制变量往往都是具有物理属性的变量，如电流、温度等。对每一个控制系统来说，控制变量 $u(t)$ 的选取是受约束的，即 $u(t)$ 的取值有一定范围。这个取值范围对应于 m 维空间中的一个集合 Ω，于是 $u(t)$ 的每一个值对应于集合 Ω 中的一个元素。Ω 称为控制域，凡属 Ω 的控制称为容许控制。在控制域外系统是不可控的，不存在其最优控制的解。

（4）性能指标。在状态空间中，要使状态向量从初始状态 $x(t_0)$ 转移到终端状态 $x(t_f)$，可以通过不同的控制函数（规律）来实现。为了从中找出一种符合人们需要的效果的最佳控制，就必须设定评价函数，它是评价控制好坏的目标，所以又称为目标函数。随着实际问题的不同要求，性能指标的选取也是不同的，性能指标 J 是控制变量 u 的函数，故可记作 $J(u)$。性能指标一般是一个泛函，最优控制实际上即是求泛函的极值。

所谓最优控制的数学描述，就是用严格的数学语言，将通常的最优控制问题抽象为一个数学问题。即寻找容许的控制作用 $u(t)$，在它作用下，使状态方程

$$\dot{x}(t)=f(x(t),u(t),t) \tag{1}$$

描述的受控对象，从已知的初始条件

$$x(t_0)=x_0, \tag{2}$$

转移到终端

$$x(t_f)=x_f, \tag{3}$$

保证性能指标取最小值，即 $\min J(u)$，找出的控制作用 $u(t)$ 称最优控制，记为 $u^*(t)$ 或 u^*。在最优控制 u^* 的作用下，状态变化过程称为最优状态轨迹，它是状态方程（1）满足方程（2）、（3）及性能指标为最小条件的解。上述问题通常称为最优控制问题，简称最优控制，对应的系统称为最优控制系统。

求解最优控制问题的实质在于以动态系统的状态方程为约束条件，求性能指标泛函 $J(u)$ 的极值问题。求解这种问题，目前常用的方法有如下几种：（1）古典变分法。（2）最小值原理。（3）动态规划法。（4）各种数值计算方法——直接法。由于计算机的发展，常需要采用数值计算方法计算最优控制作用 u^*。这种直接通过数值计算求解最优控制问题的方法称为最优控制直接数值计算方法，简称直接法，它又分为无约束和有约束的最优控制数值计算。前者有离散动态规划法、梯度法以及由此而派生的共轭梯度法、二次变分法、变尺

度法等；后者主要有广义梯度法、罚函数法等。

最优控制问题的解存在的前提条件是该系统为完全可控的，若采用反馈来实现最优控制，并用输出向量来观测状态，则系统还必须是可观的。

求解最优控制问题，通常是找出最优控制的必要及充分条件，再去求解，但由必要条件求解的不一定是最优解；根据充分条件求得的解一定是最优解，但可能不包括所有最优解。必要条件的解若满足充分条件，一定是最优解。如果已知最优控制问题有唯一解，则必要条件求得的解一定是最优解。□

4 随机系统的最优控制*

一 随机过程

在自然界、日常生活和工程技术中，随机过程与随机控制的例子不胜枚举。例如在一个电路中某一点的电流，如果此电路的运行受随机因素干扰，则对每一个瞬时，该点的电流强度都可看做是一个随机变量。从时间变量的角度，可记作 $Z(t)$，称为一个随机过程，它是时间 t 的函数。对每一个固定的时刻，它都是一个随机变量（或随机向量）。凡是受随机因素干扰的量，从时间演变的角度，都可看做是随机过程。当飞机从一个城市飞往另一城市的过程中，因风向、风速的变化而产生了导航问题，这时驾驶员在不可预测的扰动作用下进行操作。实际上，驾驶员与大自然构成一个随机系统，飞机飞行的过程就是一个随机过程，驾驶员对飞机的操纵问题，是一个随机控制问题。

一般来说，当系统具有随机输入、随机干扰或随机特性（参数）时，则系统的状态、输出和控制量也将是随机过程，这类系统就称为随机系统。对这种具有随机性系统的控制问题，就叫做随机控制问题。随机控制问题是随机问题的一个子问题，是工程控制技术中经常遇到并需要解决的复杂问题。显然，通常的确定性系统的控制理论和方法不能简单地推广到随机控制系统中来。因此，随机控制系

* 本文原载于《自动化博览》1992 年第 4 期。

统理论的研究和建立，就成为一个十分重要的问题。

设 t 是一个连续变化的变量的随机过程，称为连续随机过程。当 t 取离散值，例如 $t=\cdots,-1,0,1,\cdots$ 或 $t=0,1,2,\cdots$ 时，这类随机过程称为随机时间序列或随机序列。若根据统计规律特性，随机过程又可分平稳的和非平稳的两类。若一个随机过程的统计性质不随时间的平移而变化，或者说，不随计时起点的选取而变化，则称此随机过程为平稳随机过程；若随机过程的统计性质明显地依赖于时间的平移而变化，或随计时起点的选取而变化，则称为非平稳随机过程。在控制工程问题中，严格说来，客观中不存在真正的平稳过程，但许多实际情况，在观测期间，过程不会有明显变化，因而可以假定过程是平稳的。对于平稳过程的研究，已有“有效”的数学工具。分析系统在平稳随机输入作用下的性能比非平稳的要方便得多，所以在控制工程中最常用的是平稳随机过程。

二　随机控制的形成、发展与滤波

随机控制的形成与发展和随机系统在第二次世界大战期间军事应用的许多问题有关。早在 40 年代左右，美国麻省理工学院的学者们就用随机理论来综合火力控制系统，在詹姆斯、尼可尔斯和菲尔普斯所写的书中，采用参数最优化描写了一个有趣的例子，即跟踪雷达的设计。1940 年，美国学者维纳和前苏联学者柯尔莫格洛夫开创的滤波和预测理论，为从具有随机干扰过的观测数据中获得有用的真实状态信息提供了理论依据，这是随机控制理论发展的里程碑，在解决随机最优控制问题中具有非常重要的作用。然而，维纳、柯尔莫格洛夫方法局限于处理平稳随机过程，并只能提供平稳态的最优估值；此外，还要对维纳—霍夫方程求解，该方程是一个积分方程，它在实际应用中很少有解析解，也很难对它求数值解。由于以上那些原因，使得维纳、柯尔莫格洛夫理论没有得到广泛应用。1960 年，美国学者卡尔曼和布西提出最优滤波方法，称之为卡尔曼滤波。在这一滤波方法中，考虑了被估量和观测值的统计特性，可利用计算机来实现，从而避免了解积分方程的困难。同时，卡尔曼滤波既适用于平稳随机过程，又适用于非平稳随机过程。因此，卡尔曼方法一经出现，就在航空航天技术及其控制中得到了应用，在工业控制方面也正在获得推广和应用。

所谓滤波，主要是指从被噪声干扰过的数据中得出有用的真实信息的操作。针对不同系统所产生的数据也就有其相应的滤波方法。卡尔曼滤波器，按其实质只是一个最优递推数据处理算法。所谓滤波器，事实上只是中央处理器中的计算机程序。

随机控制理论的发展与控制理论的其他分支的发展密切相关，而随机控制理论又是其他复杂系统（如自适应控制系统、自学习理论等）的基础，所以70年代以来，随机控制及与其相伴的滤波问题，诸如随机最优控制、最优滤波以及非线性随机控制和滤波问题等，都获得了很大发展。现仅就其中的随机最优控制作简要的介绍。

三　随机最优控制

我们知道，一般的最优控制，扼要地说就是找出允许控制的方案（策略），使动态系统（被控对象）从给定的初始状态转移到某种要求的终端状态，并且保证要求的某一性能指标达到最小（或最大）。那么，什么是随机最优控制呢？我国已故的数学家关肇直教授生前给出了它的定义：被控对象上附加了随机的系统噪音和测量噪音情况下的最佳控制问题，称为随机系统的最佳控制问题。因此，随机最优控制既涉及最优控制，又涉及滤波，而且滤波是控制的先决条件。因为如果对被控系统的现实状态都不能较准确地估计，那就更谈不上对它进行控制了。为了定量地研究系统的随机最优控制，首先应该建立带有随机输入或随机干扰的系统数学模型，并在此基础上分析系统的性能。由于随机输入或干扰是白噪声过程时，系统的数学模型比较典型，分析时易于处理，因此当系统输入或干扰为其他过程时，常使其“白噪声化”。白噪声是一种理想化的模型，自然界或实际物理系统并不存在真正的白噪声过程，但是实际出现的许多重要噪声过程确实可以用白噪声过程来近似。这样，一方面既可以应用有效的、方便的数学方法来处理；另一方面又不会带来显著的处理误差。故白噪声过程在随机分析中有着特殊的重要性。以下所说过程都是指白噪声过程。

与确定性模型相比，随机性模型的建立更为复杂，其困难在于如何描述随机因素，以及如何把它在系统的数学方程中表达出来。为了解决这些问题，首先就要对随机变量有确切的数学描述，从而要知道系统的初始状态与量测噪音的统计特性，而这本身又要通过试验或观测来获得。

对于确定性系统，动态系统（被控对象）的数学模型可以是微分（或差分）方程，或状态变量表达式；对于线性或线性化系统来说，还可以是脉冲响应函数（或响应矩阵）或传递函数（或传递矩阵），它们是根据系统中变量服从的自然定律（物理的或化学的等）经过某些处理推演而获得的。对于不确定性的随机系统，当其具有随机输入、随机干扰噪声时，仍可应用确定性系统的数学模型形式，不过现在应在原有模型的基础上引入不确定性因素，加入随机输入信号，或

用随机输入（干扰）来代替确定性数学模型中的确定性输入项。对于离散系统模型的建立还比较简单，但对于连续系统模型的建立，就必须引人随机微分方程和随机积分的概念。

随机最优控制的解在很大程度上依靠动态规划的概念与方法。对于线性二次高斯系统，即被控系统是线性的，性能指标为二次型的，随机干扰（噪声）是高斯型（正态型）的二次型随机最优控制（有时称为线性二次高斯问题，或 LQG 问题，或者称为线性随机调节器问题），不仅理论上发展得比较完善，还有了比较成熟的处理方法，而且在实际上也获得了广泛的应用。它在计算机上实现最优控制时，可以分别设计为独立的两部分：最优估值程序和最优控制程序。换言之，即分两步走：① 先作滤波，得出系统状态的最优估值；② 反馈（控制），以所得估值当作系统状态的真值，再按确定性系统求得最优控制。这种分两步走的做法的根据就是分离原理。这个原理在理论上已有严格的证明，它是随机线性最优控制的一个最重要的原理。一个线性调节器问题的最优控制器的设计可转化成两个各自独立的最优化问题，在设计确定性线性控制器时不必考虑系统中所有的统计参数，而最优滤波器与性能指标中的所有矩阵无关，这样就使系统的设计大大简化，在实用上也提供了方便，故有时也将分离原理称为确定性等效原理。由于这种带有随机干扰的系统（即 LQG 问题）是真实控制系统的更确切描述，所以线性系统理论在实际中最起作用的并非确定性线性系统，而是这种线性二次高斯系统的理论。

估计问题与控制问题常伴随在一起形成对偶，即最优估值和线性最优控制问题是对偶的，如果其中一个问题解决了，我们就可借助于对偶性很容易地得到另一个问题的解。分离原理与对偶原理是随机控制系统中的两个重要理论结果。

其他线性随机最优控制问题和非线性随机最优控制问题，其中很多问题是 LQG 问题的扩展。□

5 准最优控制*

最优控制，概括地说，就是在动态系统（被控对象）的初始条件和参数已知的情况下，从所有可能的控制规律中，寻求一种最优的控制规律，使某一要求的性能指标达到最小（或最大）。而所谓最优控制系统，是指在一定具体条件下，系统的某种性能指标具有最优值的控制系统。

最优控制系统的性能，与什么样的性能指标取最优值有着密切的关系。当选取最优值的性能指标不同时，则控制系统的参数、结构以至它的类型亦有所不同。因此，控制系统的最优化问题，应包括性能指标的合理选择以及最优控制系统的设计。而在两者中，选择怎样的性能指标，又是实现最优控制的关键。因为性能指标也就是最优控制所要求达到的目标，在很大程度上决定了最优控制的性能以及最优控制是什么样的形式（如是线性的还是非线性的，是定常的还是时变的等）。为求得动态系统的最优控制规律，使性能指标达到最小（或最大），在数学上，目前主要有变分法、庞特里亚金的最小（大）值原理、贝尔曼的动态规则，此外，还有所谓数值解法（如最速下降法、黄金分割法、共轭梯度法、罚函数法等）。因此，用数学方法求解最优控制问题，一般是首先建立动态系统（被控对象）的数学模型，也就是要用动力学方程组或状态方程来描述动态系统。

最优控制虽是控制系统的一项必要的要求，但是对一些复杂多变的、大规模的（特别是多目标）系统来说，要求

---

*本文完成于 1994 年。

实现最优控制是不容易的，甚至是达不到的。但是在某些情况下，实现近似的或满意的最优控制不仅是必要的，而且是可能的。这样近似的或满意的最优控制，叫做准最优控制或次最优控制。它是最优控制系统中的一个特殊类，在某种意义下很类似于美国经济学家西蒙教授在决策科学中提出的“满意原则”。西蒙认为人的头脑里能思考和解答复杂问题的容量，同问题本身的规模相比是非常渺小的。因此，现实世界里，客观的合理举动，或者即使接近这种客观的合理性，也是难以达到的。故此，对决策者而言，最优化几乎是不可能的。西蒙认为用“令人满意”一词来代替最优化似乎更切合典型的步骤，许多实际决策行为都证实了这一点。

那么为什么真实的控制系统只是实现了准最优控制呢？概括起来有以下几个原因：

第一，数学模型只是近似的。最优控制是设计控制系统的一种方法，在这里的所谓“最优”是相对的，是在一定条件下的最优，它与实际生产中的“最优控制”的含义是有差别的。在实际生产中所谓“最优”，通常是泛指使生产控制在最好的状况上，而这个“最好的状况”往往很难用一个目标函数来刻画，缺乏一个精确的数学含义。假若状态方程能完全精确地描述生产过程，而目标函数又能完全反映生产状况的好坏，那么，这里所说的最优控制与实际生产中的所谓最优控制就一致了。可是在实际中，往往因为生产过程复杂，因素众多，在建立数学模型时，首先要把次要的、无关紧要的因素（如小惯性）忽略掉，进行简化；其次还要进行线性化（如间隙、死区、干摩擦等非线性进行线性化）；另外有些参数要由实验测得，必然带有不同程度的误差。一般对于较高阶次的最优控制系统，用变分法、最小值原理或动态规划求解最优控制时，常常先把描述过程的高阶微分方程（或代数方程）化成二阶、三阶微分方程组来描述。这样，高阶系统就简化成低阶系统。

降低系统阶次，有这样一种方法，即对于含有小参数的系统，可表示成两个子系统，其中一个含有小参数的子系统，另一个不含。也就是把一个大系统写成一个包括状态变量的状态方程组，在这个方程组中忽略小参数的影响，于是高阶系统就简化为低阶系统。因此，状态方程只是对生产过程的一个粗略的描述，或者说，数学模型是近似的，而目标函数只近似地反映控制质量的好坏。在这种情况下，从理论上得到的最优控制，只是近似地反映了生产实际情况，将它用于实际时，往往要根据实际情况作适当修改，以便达到预期的效果。事实上，还有的生产过程很难用一个数学模型去描述它，甚至提不出一个目标函数去衡量控制系统的好坏。那么，对于这种情形就不必采用最优控制这一方法去处理，而另想别法，如用计算机进行实时控制。

第二，控制规律进行了某种简化。控制规律和系统的参数、终了状态等有关，如果最优控制作用的“控制段”太多，那么，相对于系统状态的变化状态过程而言，每“段”的时间间隔太小，系统的状态来不及改变，最优控制很难完全实现。因此就需要简化控制规律，即或者预先去掉某些小的“控制段”，或者改变“控制段”的长度，或者改变控制规律等。对于多段的时间最优控制，关于其开关面方程，要求用非线性变换器实现切换面。系统的阶次越高，实现切换面要求用数量更多更复杂的非线性变换器。对于这类控制，最简单的简化方法是用线性变换器代替非线性的变换器，开关线可用近似的曲线去实现，这即是准最优控制。

关于二次型性能指标最优控制，实现最优控制要求实现全状态反馈，但在一般情况下，系统的输出变量少于状态变量，或者有些输出变量不能直接测得。在此情况下，为实现最优控制，可采用状态观测器，利用所谓输出信息重构出全部状态变量，实现状态反馈。另外，也可以只用输出变量直接实现线性反馈，这时因为输出变量数目比状态变量少，所以实现的是准最优控制。

第三，控制元件或装置非理想化。最优控制元件或装置不能实现理想特性，也是真实的控制系统只是实现近似的最优控制的原因之一。例如，实现切换面需要理想的继电器，但实际上继电特性元件有非灵敏区、磁带和惯性，不能实现理想的继电特性。因此，用真实的元件实现最优控制就会偏离计算的理论值。另外，还有其他控制元件也同样会有非理想化问题。因此，实际最优控制系统往往多是准最优控制。

最优控制理论的学科体系日趋发展，并且已有一些出色的应用，但最优控制理论在很多方面与实际应用还有距离。当理论上的最优控制无法实现时，我们可以降低一些对指标的要求，实现准最优控制。□

6 计算机实时控制*

古典控制理论本质上是一种频域法，要靠各个频率分量描述信号，这就是说只限于线性定常系统才能用频域法，否则就不能用叠加原理进行分析。此外，古典控制理论是以传递函数方法为基础的，只适用于单输入单输出线性定常系统的分析与设计。但随着科学技术的发展，特别是空间技术的发展，出现了许多大型复杂系统的控制问题，譬如火箭和导弹的特点是速度高，而在飞行中必定会碰上各种各样的干扰（如各种异常气流等），会使火箭偏离原来的既定轨道，如不及时把它纠正过来，就会越偏越远。火箭、导弹或卫星偏离轨道是时有发生的，问题是要把它及时纠正，以它的尾舵转一个角度，以便让它恰当地回到既定的轨道。考虑到火箭的速度每小时达数万千米，纠偏的命令必须迅速下达，这就要用计算机进行实时控制。在航空方面，有些航空港非常繁忙，一天有上千架飞机起降。空中虽是立体的，但有时也难免太拥挤而产生互撞情况，通常一个控制塔大约可管理 5～10 架飞机的起降。使用计算机进行实时控制，可以提高到同时管理约 250 架飞机的起降，即提高效率 25～50 倍。在一个控制中心的显示屏上可以把机场四周 100 千米内的各架飞机情况清楚地显示出来，并且还可以根据需要指示跟踪某架飞机，自动显示其航班号、高度、速度、坐标等。

对于多输入多输出系统，非线性系统及参数时变系统等的分析与研究，对控制系统提出了高速、高精度甚至对

* 本文原载于《自动化博览》1992 年第 1 期。

某性能是最优的要求。对这些新控制要求，古典控制理论是无法解决的，于是出现了现代控制理论。“实时控制”是现代控制理论是一个重要课题。

所谓“实时控制”（又称实时自动控制或过程控制），就是及时搜集检测数据，在时间域内利用计算机作为自动控制系统中的一个信息处理环节，按最优值进行自动控制或调节控制对象的一种控制方式，是对复杂系统实行自动化的一种重要手段。与用计算机进行科技计算和大量信息（或数据）处理相比较，计算机用于实时控制的明显特点是，在自动控制中，人完全不参加控制过程。在前两种应用中，计算机计算出结果，通常由人审定后去执行，而实时控制则是由计算机作出决定，并由它执行这个决定。

计算机为什么能进行实时控制呢？因为计算机里有一台电子时钟，它是由逻辑线路构成的，可以准确地告诉你某天某时多少分多少秒，甚至多少微秒，要比日常生活中用的钟表精确几十倍。时钟，众所周知，本质上是一种计时的装置。电信号是计算机的“血液”，就是一串串瞬时变化的电流和电压的脉冲，机器里有一个叫脉冲发生器的装置，用以控制整个系统的“步伐”和衡量脉冲的变化。有了在固定间隔时间发出的脉冲，也就有了频率，即每秒多少次。有了频率就有了周期，而周期又是频率的倒数，即每一次的时间是多少。因此有了固定的周期，我们就有了计时的基础，也就能够准确地计算时间了。实时控制的实质就是，先将计算机的时钟与当地的标准时间取得一致，也就是“对表”，然后根据实际需要来实现对各种仪器设备的控制过程。通常把实时控制所采用的电子计算机称为专用控制机，包括军用控制机、工业控制机。

计算机实时控制的问题是计算机最基本的应用，最早应用于火箭、导弹的制导。在工业过程控制中，应用得比较成功的是交通、化工、炼油、冶金、电力、机械等部门，如冶金工业的转炉炼钢自动化，使优质钢的合格率达到 90%以上。

这些控制过程是怎样进行的呢？第一是建立数学模型。对于人的要求来说，主要是建立能正确反映实际情况的数学模型，这需要许多方面的专家参加。如火箭制导就需要火箭专家、空气动力学家、数学家、电子学家、自动控制学专家等的共同努力，反复实践，不断修改，使之逐步完善。现代数学，如泛函分析、现代代数等为发展现代控制理论提供了分析的工具。20 世纪 50 年代后期，贝尔曼等人提出使用状态空间法，即在建立数学模型时，首先将系统的一阶微分方程组写成状态方程的形式，在此基础上再进行各种研究。1960 年，卡尔曼在控制系统中成功地应用了状态空间法，并提出了可控性与可观测性的新概念。庞特里亚金提出了最小（大）值原理，于是开始形成并迅速发展起来了以状态空间法为基础的现代控制理论。由于采用了状态空间法，这就为在时间域内对各种诸如非线性系统、时变系统、多变量系统的控制提供了有效的工具，并且便于

实现最优控制和在整个过程内进行实时控制。第二，计算机在实时控制中起决定性作用。数学模型要存储在计算机内，成千上万个数据和变量的方程要靠计算机进行处理或运算，才能把火箭偏移换算成应当采取的转动尾舵的相应角度。从前面谈到的例子不难看出，在实时过程控制中，计算机必须非常可靠，如失灵几分钟甚至几秒钟，都会导致巨大的损失，若火箭内乘坐宇航员，则后果就不堪设想，因此必须保证计算机的各个部件、元件以及整机的可靠。一些用于宇航控制或导弹控制的计算机的可靠程度，达到连续进行五万小时以上不出问题，也就是连续开机六年，中间不出一次故障。为做到万无一失，一般在实时控制过程的系统中采用双机或多机系统，其中一台运转，其余备用，一旦出了什么故障，系统立即切换，由备用机继续运行，保证整个控制不间断。

采用电子计算机进行实时自动控制，对于自动化控制系统具有重大意义。这是因为电子计算机用于生产过程中，除起到实时控制的作用外，还能及时发现事故，进行预警。一般而言，所谓生产技术事故，无非是控制对象中一些相关部分发生新的矛盾，而本身又不能及时调节恢复平衡，从而造成事故发生。由于电子计算机运算速度快，能够及时地处理大批量数据，给实时控制复杂的生产过程提供了有效的工具，不仅能及时发现事故，还可在事故孕育过程中有及早发现的可能，然后进行预警，因此也就有可能将事故消灭在萌芽状态，从而避免造成损失。所以用于实时控制的计算机，其本身要具备可靠性高的起码条件。

随着科学技术和工业生产的发展，以电子计算机为基础的生产全盘自动化的前景是诱人的。因为这不仅可极大地提高劳动生产率，且可改变工人从属于机器的地位，成为驾驭和维护自动生产线的主人。如我国鞍钢等比较复杂的大型钢铁联合企业，从采矿、选矿、烧结、炼铁到炼钢和轧钢等，由于产品质量受到很多相关因素的影响，用人工方法来控制生产是困难的，在时间上也是来不及的，若采用电子计算机进行巡回检测、自动记录数据、自动报警、直接控制生产过程，可实现最优控制。再如大型电站和电网、远距离输油管道、人造通讯卫星以及宇宙飞船的飞行轨道等，都要靠电子计算机作为控制中心进行实时控制，实现系统工程自动化。由此可见，电子计算机用于实时控制是实现自动化的中心环节，不仅可以降低消耗，而且可以提高产品数量和质量。□

7 计算机控制及应用*

一 自动化和计算机

早期的自动化采用机械、电力或液压传动的方法，在一定程度上解决了单机和单一生产线的自动问题，但这种方法自动化水平不高，只能称为机械化，不属于现代自动化的范畴。高水平自动化的主要标志是不仅减轻人的体力劳动，而且还能减轻人的脑力劳动。随着自动化科学技术的发展，自动控制已从“经典控制理论”发展为“现代控制理论”，从单变量控制发展为多变量控制，从自动调节发展到最优控制，这些都要用计算机作为高级的自动机器，也就出现了计算机控制。由于计算机具有高速运算、信息存储、计算精确、数据处理、逻辑判断、自动运行等特点，随着计算机的发展，已逐渐形成了一个“软硬兼施”的、比较完善的计算机控制系统。计算机控制系统是强调计算机作为控制系统的一个重要组成部分。在计算机控制系统中，由数字计算机代替自动控制的常规设备，对动态系统进行调节和控制，这是自动控制系统的一个革新。说得具体些，计算机控制就是把控制对象或过程的有关参数（如温度、压力、转速、飞行器的姿态、位置、速度等）进行采样，并通过输入通道进行模数变换（A/D），也可直接将数字量输入到计算机，计算机根据这些数字信息，按预定的控制

* 本文原载于《自动化博览》1997 年第 1 期。

规律（数学模型）进行计算，并通过输出通道把计算结果通过数模变换（D/A），或直接以数字量输出，去控制对象，使被控制量达到预期的目标。

二 计算机控制系统的分类

对于计算机控制系统可按不同原则进行分类。如按调节规律分类主要有：① PID控制，即比例－积分－微分控制。② 前馈控制，即在测得干扰量的大小以后，经过适当延时，就在干扰点的前方加入一个控制作用，使它正好抵消干扰对被控制量的影响，而不是等干扰产生影响后才反过来产生抑制干扰的控制作用。③ 最优控制，即是选择最优控制规律，使系统在某种工作条件和物理条件的限制下，使某种性能指标取最大值或最小值。④ 自适应控制，即当工作条件和物理条件变化时，能自动改变控制系统的特征来实现最优控制。⑤ 自学习系统，指的是系统本身能不断地积累经验，自动地改变控制器的结构与参数，使控制效果越来越好。在上述的计算机控制系统中，最优控制、自适应控制和自学习系统是比较复杂的控制系统，在系统的工作过程中需进行复杂的实时计算，所以必须采用高效的算法和高性能的计算机。

控制系统按计算机与设备的关系可分为：① 离线控制。计算机不直接作为控制系统的一个组成部分。② 实时控制，就是能够及时地收集观测数据，按最优方案进行实时处理或对受控对象进行自动控制。这是实现生产过程自动化的重要手段。在当前的新技术革命中，实时控制几乎已进入尖端技术和工业生产的各个领域之中。

控制系统还可以按照计算机参与控制的方式分类，主要有：① 计算机监控系统，简称 SCC，是由计算机对多个被控物理量进行巡回检测，并由计算机按照一定的数学模型计算出最优给定值，并传给常规调节器，由它控制被控对象。② 直接数字控制系统，简称 DDC，是由计算机对多个被控物理量进行巡回检测，并根据规定的数学模型进行运算，然后发出控制信号，直接控制被控对象。在这种控制系统中，计算机代替了常规调节器。③ 计算机网。以一台巨型计算机为中心，配有各种类型的计算机和若干终端设备，通过通讯联系构成庞大的计算机网，它不仅便于通讯联系，而且可以共享硬件与软件资源，经济、灵活、可靠，充分利用了设备的功能与其他资源。

三　计算机控制系统的发展概况与展望

控制系统引进计算机，就可以充分运用计算机强大的计算、逻辑判断和记忆等信息加工能力，开启了人类智力解放的新时代。在20世纪70年代初，微型计算机异军突起，表明计算机的发展和应用进入了一个新的阶段。由于计算机控制具有控制灵活、实现的功能多而强、精度高等特点，因此它将逐渐取代模拟控制而成为系统控制的主要形式。

追溯计算机发展的初期——20世纪50年代，便有人想将计算机用于航天或航空系统的控制，然而，由于那时的计算机体积大，消耗功率多，可靠性不高，因此这样的想法在当时不能变成现实。到50年代末，已经有计算机控制系统在工业生产过程中投入运行，从此，计算机控制技术得到了迅速的发展，但主要是应用在炼钢、化工及发电等部门，其主要任务是寻找最优的运行条件，进行生产的管理和调度。由于早期的计算机控制属管理控制、操作指导控制或设定点控制，故仍需要常规的模拟控制装置。1962年，英国帝国化学工业公司利用计算机控制完全代替了原来的模拟控制，因而称这样的计算机控制为直接数字控制(direct digital control，简称DDC)。

我国著名科学家钱学森教授在其巨著《工程控制论》(修订本)序里写道："高精度、高可靠性自动调节、自动控制和自动监测系统，对核能技术的发展极具重要性。"在核电站发展的早期，一般采用常规的机电自动控制技术和仪表。1963年，在核电站调节、控制与监测工作中首次引用了电子计算机控制，并获得了很大成功。到60年代末期，电子计算机控制已在核电站上广泛应用，全面采用电子计算机监视和控制，是当前核电站技术发展的显著特征。计算机控制之重要，由此可见一斑。

1972年诞生的微型计算机，其最突出的优点是体积小，速度快，可靠性大大提高，而成本却大大降低，因而容易普及和推广，这就使得不管多么小的控制任务都可由微型计算机完成。同时，微型计算机的出现也使计算机控制系统的结构形式发生了根本性的变化。过去，由于计算机价格昂贵，一台计算机必须完成很多控制任务，因而采用集中式的控制结构。现在由于微型计算机价格便宜，又具有上述优点，因而改用多微处理机的"分散控制"。由于它的每个关键部位都可以考虑冗余措施，保证在发生故障时不会造成停产检修的严重后果，所以70年代中期以后分散型结构在计算机设计和工业用仪表系统设计中得到极大的重视。由微型计算机控制的单机自动化和多层的中央处理机控制的多

机自动化以及无人车间的出现，极大地提高了劳动生产率和产品质量。

除了在过程控制方面计算机控制日趋完善，在机械控制、航天技术和各种军事装备中，计算机控制也日趋成熟。例如，通讯卫星的姿态控制、卫星跟踪天线的控制、电气传动装置的计算机控制、计算机数控机床、工业机器人、飞行器自动驾驶仪等。在工业机器人的控制中，不使用计算机控制是无法完成控制任务的。国外最近研制成功由计算机(微型)控制的无人驾驶自动汽车，人们把它叫做智能汽车。当它沿着公路急驶，如果前面突然有一个人横穿过来，或者临时发现前面有什么障碍物，汽车立即自动刹车，而当路人已经穿过，或者障碍物移走，它又徐徐开动，向前驶去。人们利用计算机控制可以完成常规控制技术所无法完成的任务，可以达到常规控制所不能达到的高性能指标。

最后，我们用控制论学家宋健教授的话来结束本文："计算机的广泛应用对控制论的发展具有划时代的意义，然而，作为普遍的理论，仍处于探索阶段。"□

8 自动控制系统的稳定性*

任何一个自动控制系统，首先必须要稳定，因为不稳定的系统不能正常工作。所以稳定性是控制系统的一个重要特性，研究系统的稳定性一向是控制理论的首要问题。

什么叫系统的稳定性？若系统具有平衡状态，当外界对系统的干扰使系统偏离平衡状态时，系统是否能消除干扰恢复原平衡状态，或以什么样的精度恢复原状态的性能，这就是所谓系统的稳定性问题。因为控制系统是一种动态系统，当外界环境对它没有输入时，它处于平衡的、稳定的状态；一旦外界给以输入（包括干扰在内，干扰也是一种输入）时，它就失去平衡，处于不稳定的状态。为了再达到稳定状态，就要经过一个过渡过程，控制系统必须在这个过渡过程中才能实现稳定。这个过渡过程在自动控制理论中称为动态过程。

稳定性理论是自动控制理论、工程控制论中的一个重要组成部分，它是建立在数学的稳定性理论基础上的。在历史上，它早就引起了数学家和力学家的兴趣，并且成为这些领域内推动科学发展的一个激励因素。19 世纪末期，法国杰出的数学家庞加莱在其《微分方程所定义的积分曲线》一书中首先用定性的方法研究微分方程解的稳定性问题。1892 年，俄国数学家李亚普诺夫从数学理论方面给出了研究稳定性的普遍方法。可是因为在相当长的一段时间内，他研究的是关于线性定常系统的稳定性，所以李亚普诺夫方法在当时条件下并不实用，也没有引起人们的重

* 本文原载于《自动化博览》1992 年第 3 期。

视。在工程上最实用的是和李亚普诺夫同年代出现的对于单输入、单输出线性系统或定常系统的罗斯和胡尔维茨的代数法，以及 1932 年美国电信工程师奈奎斯特提出的频率法。因为切合工程实际，所以这两种方法直到现在仍为控制工程师所喜用。可是这些方法只适用于研究线性定常系统的稳定性，不适用于非线性或时变系统，而在现实的控制中，人们遇到的自动控制系统中几乎都是非线性的。因此非线性系统的稳定性研究，不仅在理论上极为重要，也有助于实际问题的解决。

古典控制理论认为，线性系统和非线性系统的稳定性不大一样。一个线性系统是否稳定，和系统的初始状态及外界扰动的大小都没有关系；非线性系统则不然，其稳定性和这些因素都有关系。因此，古典理论中没有给出过稳定性的一般定义。李亚普诺夫于 1892 年从数学理论上给出的对于任何系统都适用的一般化的稳定性定义和两种判别方法，随着近代控制理论和计算技术的发展，特别是计算机的出现和应用，又重新受到了重视，并得到进一步的发展。它在近代控制论中的最优控制、最优估值和自适应控制等方面都扮演着重要角色。

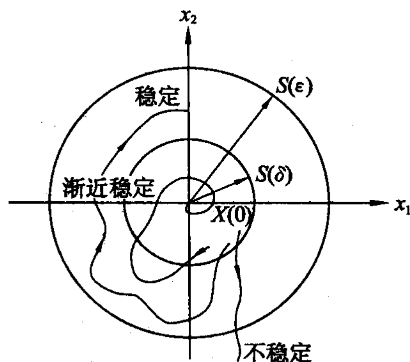
李亚普诺夫首先根据系统的输出（在控制理论中也称为响应）是否有界来定义稳定性的。因为任一系统的稳定性问题，实际上就是扰动对于运动的影响问题。只要这种影响不超过一定的界限，就可看做是稳定的。他根据这个标准区分了三种情况：

（1）稳定的。对于系统初始值的一个扰动，如果系统状态响应的幅值是有界的，就称系统是稳定的。

（2）渐近稳定的。对于系统初始值的一个扰动，如果系统状态的响应能够最终回到原来的平衡状态，则称该系统是渐近稳定的。

（3）不稳定的。对于系统初始值的一个扰动，如果其响应的幅值不是有界的（即无界），则称该系统是不稳定的。

下面我们以二阶常微分方程所描述的二阶控制系统为例，来说明系统的稳定、渐近稳定和不稳定的概念。这三种情况可以用一个平面图形来表示（如图）。



二阶系统三种情况的图示

对于以任意给定的正数 ε 为半径所作的圆 $S(\varepsilon)$，若能找到一个适当的以 $\delta>0$ 为半径所作的圆 $S(\delta)$，当时间 t 无限增大时，从 $S(\delta)$ 内发出的状态轨迹总是不

能离开 $S(\varepsilon)$ 圆内，则这个二阶系统就称为在李亚普诺夫意义下是稳定的，或简单地称为稳定。假若当 t 无限增大时，由 $S(\delta)$ 出发的轨迹，不但不离开 $S(\varepsilon)$ 圆，而且最后收敛于原点（平衡点）$X(0)$，则对应的系统就是渐近稳定的。假若有某个 $\varepsilon>0$，则不论 $S(\delta)$ 的半径多么小，由 $S(\delta)$ 出发的轨迹总能跑到 $S(\varepsilon)$ 圆外，则这个系统就是不稳定的。

这三种情况用数学语言来描述就是：

1. 对于任意的 $\varepsilon>0$，必有 $\delta>0$，使得 $||x(0)||<\delta$ 时，就能保证 $||x(t)||<\varepsilon$，则对应的系统是稳定的。（$||x||$ 表示 n 维空间中状态向量到原点的距离，称为范数）

2. 对于任意的 $\varepsilon>0$，必有 $\delta>0$，使得当 $||x(0)||<\delta$ 时，不仅 $||x(t)||<\varepsilon$，且有 $\lim\limits_{t\to\infty}||x(t)||=0$，则称系统是渐近稳定的。

3. 对于某一 $\varepsilon>0$，无论怎样取 $\delta>0$，$||x(0)||<\delta$ 时，总能找到某个 $t_0>0$，使得 $||x(t_0)||>\varepsilon$，则称系统是不稳定的。

在以上三种情况中，经典控制论所研究的稳定只限于渐近稳定，而把另两种情况看做是不稳定的。可见，李亚普诺夫的稳定性概念具有一般性，他概括了线性及非线性系统的一般情况。

不仅如此，李亚普诺夫还证明了由一般常微分方程所描述的系统可用两种方法来分析其稳定性。

李亚普诺夫第一方法（间接法）：这个方法是把非线性函数用近似级数表示，然后用近似的方法解这个非线性方程，得到特解或通解，并分析其稳定性。因为要求解微分方程，所以人们就称它是一种间接方法。它是一种定量方法。根据这一方法可以严格证明以下结论：

如果非线性方程在一个平衡点的邻域线性化形式为

$$Z=\boldsymbol{A}Z+O_2(Z),$$

其中 $O_2(Z)$ 表示 Z 元素的二次和二次以上的项，$\boldsymbol{A}$ 为合适维数的矩阵。若 $\boldsymbol{A}$ 的特征值均无零实部，则非线性系统的平衡性质与线性系统 $\dot{Z}=AZ$ 所表示的平衡性质是一样的。

李亚普诺夫第二方法（直接法）：李亚普诺夫第二方法是不直接通过求解状态方程，也不具体画出轨迹的图象，而是找出一种叫做李亚普诺夫函数的纯量函数来判别系统是否稳定。如果一个系统存在一个李亚普诺夫函数，就可以充分肯定这个系统是稳定的，而不必去求解这个系统的微分方程。由于李亚普诺夫第二方法不用解方程就能直接判别系统的稳定性，所以这种方法又称为李亚普诺夫直接方法。这是一种定性方法。对于现代控制理论来说，它具有更明显

的优越性。因为这个方法既适用于线性系统，也适用于非线性系统和时变系统。众所周知，非线性系统和时变系统的状态方程都是难以求解的，且其稳定性往往与外界条件有关，很复杂，这给非线性系统的稳定性分析带来极大的困难。用李亚普诺夫直接法不仅能判别系统是否稳定，还能分析线性或非线性的瞬时响应。但必须指出，李亚普诺夫的这个方法给出的是系统的稳定性的必要条件，未必是充分条件。正确地讲，一个系统可存在无穷多个李亚普诺夫函数，但是找不到某个特定系统的李亚普诺夫函数并不意味着这个系统是不稳定的。

李亚普诺夫直接方法有明确的物理意义，它是从能量的观点分析系统的稳定性的。如果一个系统储存的能量是逐渐衰减的，这个系统就是稳定的；反之，假若系统不断从外界吸收能量，系统的能量就越来越大，那么，这个系统就是不稳定的。如果用纯量函数 $V(x)$表示系统的能量，$V(x)$就应该是个正值。如果$\frac{\mathrm{d}V(x)}{\mathrm{d}t}$是负值，系统的能量逐渐减少，系统就是稳定的。李亚普诺夫直接法就是用 $V(x)$和$\frac{\mathrm{d}V(x)}{\mathrm{d}t}$的正负来判别系统的稳定性。也就是说，对于一个给定的系统，只要能找到一个正的函数 $V(x)$，而$\frac{\mathrm{d}V(x)}{\mathrm{d}t}$是负的，则这个系统就是稳定的。函数 $V(x)$叫做李亚普诺夫函数。

李亚普诺夫直接法是一个普遍的方法，它的最大优点是无论线性还是非线性系统都能适用。这样一来，判别系统稳定性的问题，就可以归结为寻找李亚普诺夫函数 $V(x)$的问题。过去要想寻找到李亚普诺夫函数主要是靠试探，要凭个人的经验和技巧，这也就是李亚普诺夫直接法长期以来不能推广应用的主要障碍。计算机的发展正在排除这一障碍。人们在计算机上不仅可以找到所需的李亚普诺夫函数，而且还能找到系统的稳定区域。因此，这种方法在现代控制理论中得到广泛的应用。□

9 自适应控制技术及应用*

“自适应”这个术语是从生理学借用过来的。一个细胞，一个器官，一个有机体或者一个物种，如果在变化着的环境中，能调节自身的性能以维持生理平衡，那它就是自适应的。人类社会和自然界的进化过程就是一个不断适应和发展的过程。随着科学技术的进步，仿生学异军突起，自适应控制系统在某种意义上讲，就是按照仿生学的思想而设计的一种人造控制系统。

所谓自适应，一般是指系统按照环境的变化，调整其自身，使得其行为在新的或者已经改变了的环境下，达到最好或者至少是容许的特性和功能。这种对环境变化具有适应能力的控制系统称为自适应控制系统。

由于自适应控制的对象是那些存在不定性的系统，所以这种控制应首先能在控制系统的运行过程中，通过不断地测量系统的输入、状态、输出或性能参数，逐渐了解和掌握对象，然后根据所得的过程信息，按一定的设计方法，作出控制决策去更新控制器的结构、参数或控制作用，以便在某种意义下使控制效果达到最优或次优，或达到某个预期的目标，按此设计思想而建立的控制系统便是自适应控制系统。

在 20 世纪 50 年代初，自适应控制系统便已出现，当时主要是根据设计一个高性能的飞机自动驾驶仪的要求而提出来的。当飞机以不同的速度和高度飞行时，飞机参数有很大的变化，因而要求所设计的控制系统具有适应能

---

* 本文原载于《自动化博览》1996 年第 1 期。

力，使得飞机的动力学参数在很大范围内变化时，飞行控制系统仍具有令人满意的性能。60 年代，控制理论的很多成果，例如状态空间理论和稳定性理论的结果，还有随机控制理论的重要结果，对发展自适应控制起了重要作用。贝尔曼提出的动态规划加深了人们对自适应过程的理解。系统辨识和参数估计在此期间也有重大进展。70 年代，当把不同的估计方案和各种设计结合起来时，自适应控制又重新恢复了它的活力。

在自适应控制系统中，除了受控对象与测量元件之外，一般还包括以下三个部分：① 系统性能计算（或参数估计）部分。该部分用以对受控对象进行实时在线检测，求得当前的动态特征。② 控制决策部分。根据计算（估计）提供的信息，按预定的系统性能指标，实时在线地求出所需的控制器。③ 可调控制器部分。在自适应控制系统中，控制器（结构和参数）是可调的，它由控制决策部分作在线调整，以保证随内外环境的变化而具有自适应能力。系统性能计算部分和控制器决策部分合在一起可称为自适应机构部分。

自适应控制系统的类型很多，从不同角度有不同的划分。以下给出自适应控制的主要类型，简单说明它们的基本工作原理。

1. 自校正控制系统

自校正控制系统必须连续地提供受控系统的当前状态信息（辨识），然后将系统的当前性能与期望的或最优的性能进行比较，作出使系统趋向期望的或最优的性能的决策，对控制器进行适当地修正，以驱使系统接近最优状态。

理论分析和实践表明，自校正控制技术特别适用于结构部分已知和参数未知而恒定或缓慢变化的随机受控系统。由于大多数工业对象都具有这些特征，再加上自校正控制技术理解直观，实现简单经济，所以它在工业控制中已得到广泛应用，现已成为十分重要的一类自适应控制系统。

自校正控制系统由内、外两个环路组成。内环回路由控制对象和常规的反馈控制器组成。外环回路由参数估计和控制器设计两部分组成，用来调整控制器的参数。参数估计和控制器设计必须在线地实现，因此参数估计必须采用递推算法。控制器设计必须采用计算尽量简单的设计方法，通常采用最小方差控制算法。因此，这种调节器是以最小二乘辨识和最小方差为基础的自校正调节器。如果参数辨识的算法和控制算法的选取不同，原则上可以构成各种形式的校正调节器。

2. 模型参考自适应控制系统

模型参考自适应控制系统也是一类重要的自适应控制系统。它的主要特点是实现容易，自适应速度快，并在许多领域得到了应用。

这类自适应控制系统实际是在原来的反馈控制系统的基础上再附加一个称

为参考模型的辅助系统和一个控制器参数的自动调节回路。其中参考模型的输出响应直接表示系统希望的动态响应，所以参考模型相当于输出响应的一个样板。这种用模型输出来直接表达对系统的性能要求，往往是很直观很方便的。

像自校正控制系统一样，在结构上也可把模型参考系统划分为内、外两个环路。内环路是由对象和控制器组成的常规反馈回路。外环路是调整控制器参数的自适应回路，它也是一个反馈控制回路。控制器参数通过外环回路来调整，调整的目标是使得系统的实际输出与模型输出之误差变小。

模型参考自适应控制系统的关键在于如何设计自适应控制律（自适应律），以得到一个能使误差变小并逐渐趋近于零的稳定系统。

关于自适应控制律的设计目前存在两类不同的方法，其中一类称为局部参数最优化设计方法，它利用最优化技术搜索到能满足要求的一组控制器的参数，使得某个预定的性能指标达到最小。另一种基于稳定性理论的设计方法，其基本思想是保证控制器参数的自适应过程是稳定的。由于自适应控制系统一般都是非线性的，因此，这种自适应控制律的设计自然要采用非线性的稳定理论。

3. 自寻最优控制系统

这是一种自动搜索和保持系统输出位于极值状态的控制系统。许多工业对象的极值特性在运行中都或多或少地会发生漂移，因而无法采用常规控制策略。例如，远程飞机的最优燃耗（每升汽油能飞行的最大千米数）是一个极值点，这个极值点与飞机的发动机功率和转速以及飞行环境有关。当飞机在结冰的气候条件下飞行时，机翼和机身表面冰层的堆积与溶化会使飞机外形发生变化，进而引起极值点漂移。冰层的堆积和溶化情况无法测量，因而无法预估它的影响。对于这类受控系统，采用自寻最优控制策略便可自动保持极值运行状态，使运行状态的梯度为零。此外，自寻最优控制系统还具有易于理解和实现方便等优点，所以它在工业中有广泛的应用。

显然，为保持最优极值点，自寻最优控制系统必须具备两个功能，即探测和调整。探测是为了获取当前运行状态的信息，调整是为了保证最优运行条件。实现这两个功能的主要方法有切换法、摄动法、自导法、模型定向法。

自适应控制系统的进一步发展，将走向“自学习”系统和“智能控制”系统。这类系统除具有前述一般的自适应功能外，还应当拥有大型记忆、模式识别以及各种各样带有智能性的高级决策功能。这类系统能够记住系统过去的经验和教训，识别曾经发生过的情况，并且能够基于过去的经验来逐步改进其自适应动作。设计这类系统所采用的方法除解析方法外，更重要的是模拟人类行为的方法，即智能模拟。□

10 自寻最优控制系统*

一　自寻最优控制系统的缘起及意义

经典的最优控制就是：按照被控制对象的动态特性，找出一个容许控制方案，使被控制对象按照技术要求运转，并最终使某一性能指标（而不是使任何性能指标）在某种意义下达到最优（最大或最小）值。求解最优控制的数学方法目前主要有变分法、庞特里亚金的最小（大）值原理、贝尔曼的动态规划以及各种数值解法，应用这些方法已较成功地解决了许多动态最优控制问题，如最少时间控制、最少燃料控制和最佳调节器等。最优控制理论不仅成功地应用于宇航、航空、航海等领域，还在生产设计和生产过程如化工、冶金、电力方面也获得广泛的应用。

但是，用数学方法求解最优控制问题，首先要建立被控制对象（系统）的数学模型（即描述动态变量的状态方程）。一般说来，建立某个特定系统的数学模型是一件非常困难的工作，越是复杂的系统越是如此，除了掌握有关这类系统的理论分析和大量的实验数据以外，还需要相当的实际知识和技巧。另外，一个系统总是受环境干扰而遇到一些无法测量的（随机）干扰因素，有时想建立一个完全准确的数学模型几乎是不可能的。另一方面，有时系统在运行过程中可能要改变，而这种改变又难以用已知的规律

---

* 本文原载于《自动化博览》1991 年第 6 期。

去描述。在这种背景下，美国自动控制学家椎柏(Draper)和拉宁(Laning)等人为了让设计出来的自动控制系统，能使被控对象处在最优工作状态或接近最优工作状态，设想把本来是由设计人员事先应设法了解与掌握的被控对象的性质与特征的这一任务，在控制系统运行的过程中，由控制系统本身来完成。根据这一设想，在进行控制系统设计时就不需要有关被控对象的确切知识，而是采用在运行过程中“不断测量”和“不断理解”的方法，由系统本身去摸清当时系统运行的条件，能适应过程动态的变化而修改控制动作，以保证系统能按事先的要求运行。这样一来，设备制造和运行过程中产生的与原设计的不一致，以及种种事先无法估计到的变化所引起被控对象特性的漂移，都能在系统运行过程中，通过不断测量而被察觉出来，由系统本身根据运行条件的变化不断地修改控制作用，使被控对象能在最优工作状态下运行。这种自动寻找最优工作运转点的自动控制系统，称为自寻最优点控制系统，简称自寻最优控制系统。

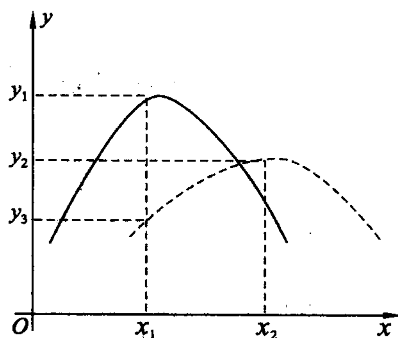
由此看出，自寻最优控制的基本思想是：让控制系统能根据变化的情况，自动改变控制量，使被控量维持最优或次优水平。一个自寻最优点的控制系统，总是力图使一台设备或生产过程达到或接近某一最优指标。

自寻最优控制是在模拟人完成工业控制的操作，它不需要知道被控对象的数学模型，因而它不仅避免了建立数学模型的困难，而且对系统中不确定因素，如随机扰动所引起对象特性的漂移等，具有自适应的能力。由于自寻最优控制具有容易理解、便于实现的优点，因而在生产中易于推广。对于要求某项性能指标达到最优的生产过程来说，自寻最优控制是一种极为有效的方法。

二　自寻最优控制的工作原理和特点

自寻最优控制的工作原理是利用被控对象的极大值、极小值或其他非线性静态特性，改变控制量，试探和自动搜索它对性能指标的影响，从而确定相应的运转条件，使某一性能指标达到或接近最优。它的特点是：① 只需要知道被控对象具有非线性(如极值)特性，而不需要关于被控对象精确的数学模型表达式，故与通常用计算机实现的最优控制不同。对于后者，则需要首先建立被控对象的精确数学模型，而这是一件非常困难的工作。② 自寻最优控制事先不知道控制量的给定值，而是系统去自动搜索最优工作点，以保证性能指标达到或接近最优，故也不同于通常的反馈控制。众所周知，反馈控制是按偏差进行控制的，即不断把控制对象的实际结果与目标值进行比较，控制的结果是使其差值尽可能减小。故反馈控制本质或特征在于使控制对象的状态经常与目标值

保持一致。③ 自寻最优控制系统所寻找的最优值(即被控量)不是固定值，也不是预先给定的，它决定于生产过程中的状态，是由自寻最优控制系统在控制过程中不断搜索(检测、计算、判断)而获得的。如图中对应最优性 y_1 的控制量为 x_1，当工况变化使对象性能静态特性产生漂移后，系统搜索到新的最优性能指标为 y_2，对应的控制量为 x_2。y_2 可能低于 y_1，但如不作修正仍保持在 x_1，在新的工况下对应的控制指标为 y_3，比 y_2 更差。因此自寻最优控制系统始终保持在最优工况条件下工作。④ 自寻最优的工作过程是一个自动搜索的过程。由于它能在变化了的环境下自动寻找最优工作点，因而它对于对象特性变化具有自适应的能力，这是它最为可贵和独特之处，是其他最优控制方式所难以达到的。



虽然自寻最优控制具有上述特点或优点，但其应用必须具备一定条件方能发挥其特点。其应用的必要条件是：① 被控对象具有单峰极值静态特性或其他非线性特性。自寻最优控制大多是利用被控对象的这种特性工作，或利用其非线性静态特性斜率的变化为依据进行搜索。② 极值特性是漂移的。自寻最优控制的突出优点是对于对象特性的漂移具有自适应的能力。如果极值特性不漂移，则可通过实验找到一个固定的控制量，使性能指标保持最优，以后稳定在这一参数就可以了。在此情况下，采用普通调节就可满足要求，因此不必采用自寻最优控制。③ 系统是可控的。自寻最优控制是通过改变控制量，试探其对被控指标的影响，作出判断以进行控制。如果系统是不可控的，则改变控制量对被控指标没有作用，或者不能获得正确的反应，或者干扰影响超过了控制量的作用，在这些情况下，系统都不能正常工作。此外，最优指标还必须是可测或可计测的，否则就不能采用自寻最优控制。因为自寻最优控制是通过观察最优指标的变化，以改变控制量的调节方向。因此，被控指标的值必须通过测量获得，或者通过测量其他参数后计算而获得。

自寻最优控制可以扼要表述为：调节被控对象的输入量，试探其对输出的影响，将表现输出偏离最优点的信号反馈到控制输入量的执行机构，改变输入量的调节方向，使输出量达到或接近最优点。为达此目的，可以有各种不同的方法，如导数判断法、相位判断法、极值记忆法、步进搜索法等。

三　自寻最优控制在工业自动化中的应用

自寻最优控制便于实现，易于推广。早在20世纪50年代末期，中国科学院自动化研究所等单位就已开始研究自寻最优控制系统问题，并运用于工业自动化上。但是，由于当时电子技术水平的限制，利用继电器、电子管、晶体管等元件制成的自寻最优控制器，可靠性低，维修困难。直到1969年，机械工业部北京工业自动化研究所根据自寻最优控制的原理，研制出了一种极值调节器，把它用于高炉热风炉燃烧过程的自寻最优控制，实现了合理燃烧，取得节约煤气、降低焦比、提高风温、延长炉龄等效益。1982年，河北省科学院自动化研究所按上述极值调节器原理研制了用CMOS集成电路组成的ZJT-I型调节器，用于邢台钢铁厂热风炉，仅煤气一项，每台热风炉一年可节约费用达5.5万元。

由此可看出，自寻最优控制已经取得了明显的经济效益。由于现在大规模集成电路技术及微型计算机的广泛应用，为自寻最优控制器的研究提供了先进的、可靠的技术工具。各种电子元件和计算机软件可以极为方便地赋予它完成人们所期望的逻辑判断、计算和各种控制功能，并且运行可靠、价格低廉。这些为实现自寻最优控制创造了极为有利的条件，为自寻最优控制技术在工业自动化中的推广应用开拓了广阔的前景。□

11 自适应、自学习和自组织控制系统*

一 自适应控制系统

人的体温在正常情况下受到机体自身调节系统的控制，保持相对的稳定。如环境温度较低，机体便自动改变体表组织和血流量，以保证身体与环境之间的最优热传导条件；如果环境温度变高，机体便借助于出汗和呼吸的过程，充分释放过剩的热量。人体在体温调节方面就是一种典型的自适应控制系统。随着物种的进化，高等生物在体内建立起越来越完善的自动控制系统。这些控制系统不仅保证了生物机体内环境的相对稳定和肢体运动的精确而有效，并且为机体的生存创造了更为有利的条件。对人体来说，内环境包括体温、血压、血液中激素的含量、体液中各种离子的浓度等，它们的相对稳定是生命存在的必要条件。至于机体如何才能保持内环境的稳定，这是生物控制论研究的重要内容。与工程控制系统一样，人体内环境的相对稳定性大多是由反馈控制来保证的。

自适应一词本来是描述人类、动物、植物按照环境的变化调整其本身特性的能力，现在已推广到工程技术中。一个系统，当其某些特性（如结构、参数等）为已知时，一般是能达到最优控制的。即使系统的输入信号与噪声（或干扰）中具有随机的因素，也可以通过状态滤波器进行滤波，

* 本文原载于《自动化博览》1990年第6期。

以达到最优控制。但事实上，系统都是与环境相互联系着的，环境的变化往往影响到系统的某些特性，而这些影响又是不能事先知道的，尤其对于复杂系统更是如此。人们在对系统实现控制的过程中，该系统的特性实际上已经发生了不同程度的变化。事先确定的最优控制在变化后的条件下，可能已不再具有最优性了。因此设计一种随着环境或其他条件的变化能自动调整有关的参数，以保证系统处于最优状态的最优控制，便成为控制论工作者研究的热门课题。自适应、自学习和自组织系统便是在这种背景下产生的三个既相互联系又相互区别的基本概念，它们在不同的学科中有不同的涵义，很难给出精确的定义。但从以上所举事例可知，自适应一般是指，当客观环境发生变化，系统能改变其自身，使得其行为在新的环境下，仍能按照外界条件的变化，自动调整其自身的结构或参数，以保持系统原有功能。这种系统就称为自适应控制系统。自适应系统在生物界是普遍存在的，如人和动物的瞳孔在黑暗中会自动放大，人们可以通过出汗而降低体温，又可以通过加速提供热量来弥补散失的热量，从而使体温保持恒定等。

自适应控制系统的出现还可追溯到20世纪四五十年代，它是指具有自适应功能的控制系统，例如一些简单的极值控制系统(如自寻优点的系统)和简单的被动自适应系统(如条件反馈系统)。一般的反馈控制系统和随机系统，在不大的外干扰作用下也具有一定的适应能力。例如调温系统能在周围环境温度变化不大的情况下，将温度保持在某一温度附近。但当周围环境在大范围内变化时，这种系统就不再具备适应能力了。而自适应系统则能够在大范围变化的环境中调整自身的特征，以便按照所希望的性能运行。60年代后，随着自动控制理论和计算机的发展，自适应控制系统也日益完善，并从航空航天等尖端技术部门逐渐发展到许多工业部门中。

自适应系统概括起来具有两个基本功能：① 能够自动地测量和分析系统所处的环境和对象的特性。这就是对环境或对象的识别问题，它与系统辨识密切相关。② 能根据从环境和对象所得的信息作出决策，改变调节规律，以保持系统的稳定和实现某种最优控制。无论识别还是作出决策，都要进行计算和推断，这可在计算机上实现。

通过关于自适应控制系统的研究，目前已出现的自适应控制系统主要有以下两类：

(1) 参考模型自适应控制系统。这是理论上比较成熟、应用上比较广泛的一类自适应控制系统，它由参考模型、被控对象、常规反馈控制器和自适应控制律这几个部分组成。这一类系统的特点是：它总包含一个模型作为系统的组成部分，这个模型称为参考模型(或理想模型)，在运行中总是力求被控过程与参

考模型的动态相一致。

设计这种控制器的关键在于如何综合自适应控制律，目前主要有三种方法：① 参数最优化方法，即利用最优化技术搜索到一组控制器的参数，使某个性能指标达到最优。优化的方法一般采用梯度法。② 利用李亚普诺夫的稳定性理论。参数优化方法是有缺点的，用它设计出来的系统不稳定，因此需要利用李亚普诺夫的稳定性理论和方法。此法的基本思想是，使参考模型的动态与实际过程尽可能一致，以保证闭环系统是稳定的。③ 利用波波夫的超稳定性理论来设计自适应控制律。李亚普诺夫的稳定性理论是在不考虑外界输入的情况下系统本身的稳定性质。在有外界输入的情况下，如何考虑系统的稳定性呢？波波夫提出的超稳定性理论就是针对这种情形的一种稳定性理论。用超稳定性理论设计参考适应模型是近几年发展起来的一种有效方法，它在电传动上得到了成功的应用，但这种方法需要传递函数为严格正实矩阵。所谓严格正实矩阵，就是传递函数的每一元素都为严格正实函数。近年来，这种自适应控制系统已被用在恒星跟踪系统的控制上及舰船的自适应驾驶仪上。

（2）自校正调节器。首先对被控制对象的参数随环境的变化情况或直接对环境的变化进行在线辨识，然后把辨识获得的信息送到可变调节器中，使调节器能适应当时的环境，以达到优良的控制性能。这种自适应控制的形式称为间接控制，它包括两项功能：① 辨识功能，即对对象或环境的变化进行辨识。② 决策功能，即根据预定的设计进行计算并改进调节器的参数。这两项功能都需要在计算机上来完成，通常采用最小方差控制算法，因此这种调节器是以最小二乘辨识和最小方差控制为基础的自校正调节器。

二　自学习控制系统

自适应系统的进一步发展，出现了自学习系统。在控制论中，学习一词被简单地理解为“通过经验取得某种技能，而不需要外界帮助”。具有学习功能的系统称为自学习系统。

一个学习系统，若其学得的信息被用来控制一个具有未知特征的过程，则称为学习控制系统。人们在设计一个自学习系统时，可以不必知道有关控制的确切算法，而只要给出达到这一算法的途径就够了。然后，系统便根据自己运转过程中的经验，不断修正与改进算法，使之达到最优或次优的程度。

在控制系统中，所谓学习控制是指系统可以根据过去的经验积累，不需要对系统进行连续识别，就可以自动地采取自适应方式，使系统运行在最佳状态。

这是一种更高级的自适应控制系统。

三 自组织控制系统

任何系统的形成都是一个组织过程，这是事物或一组变量从无联系的状态进化到某些特定状态的过程。一个系统在开始时，它的各部分是分散的，每一部分是独立运转的。在运转的过程中，系统的各部分之间形成联系，从而形成一个有组织的统一整体。这种具有自行组织功能的系统，称为自组织系统。自组织现象是普遍存在的，在物理系统和生物系统中更是屡见不鲜。普利高津的耗散结构论和哈肯的协同论都是研究一个系统如何能够自发地产生一个系统和如何从无序走向有序的，也即是研究自组织规律的。

对于那种不能完全预知其环境和控制对象，也不能确切知道其控制算法与条件，但能够通过自身的运转，逐步积累经验，改进控制，以逐步达到最优或次优的系统，称为自组织系统。对于控制论研究工作者来说，就是要研究和构造一些人工的自组织系统，这也是目前人工智能研究的一种途径。□

12 工程控制论*

工程控制论极大地推动了自动化科学技术理论的发展，已被广泛应用于各个领域。

我国杰出科学家钱学森创建的工程控制论，享誉世界，产生了深远的影响。本文扼要介绍工程控制论的含义、基本内容及其应用领域。

一 什么是工程控制论

什么是工程控制论？虽然对它的研究已有近40年的历史，然而至今仍没有一个众所公认的统一定义。20世纪40年代末50年代初，我国著名科学家钱学森在对第二次世界大战后迅速发展的控制与制导工程技术实践全面观察时，用敏锐的眼光发现、提炼出指导控制与制导系统设计的普遍性概念、原理、理论和方法，从而创建了作为一门技术科学的工程控制论。这里所谓制导，就是对一个飞行器从目前位置到将要到达的位置轨道上的导航进行控制。

钱学森在《工程控制论》修订版序言中说："工程控制论在其形成的时候，就把设计稳定和制导系统这类工程实践作为主要研究对象。"他总结了过去别人研究的成就，又加上他自己多年来研究的成果，于1954年用英文写成专著《工程控制论》并在美国出版，标志着工程控制论这门新的技术科学的诞生。该书也被公认为是工程控制论的奠

* 本文原载于《自然杂志》1993年第11、12期。

基性著作，此后，俄文版、德文版、中文版分别于1956、1957、1958年相继问世。《工程控制论》书中所阐明的基本理论和观点，除了奠定工程控制论的基础外，还指出了进一步研究这门技术科学的方向，极大地推动了自动化科学理论的发展。该书出版二三十年后，为适应我国科学技术发展的需要，应广大读者的要求，钱学森早在1963年就委托控制论专家宋健等人对《工程控制论》原书进行补充修订，修订本于1980年问世，它在保留原书的几乎全部内容的基础上，按原书的宗旨，选择新的材料，增补了近20多年来这门学科发展的新成果，使工程控制论的内容更为丰富和完善。

工程控制论没有公认的统一定义，我们从其历史发展过程加以考察：在控制论出现以前，即1945年前后，已形成了以伺服机构理论为核心的自动控制理论；在1948年，维纳发表了著名的《控制论》，诞生了控制论这一新兴学科。这两方面的理论相互渗透、结合，使得控制论的基本理论和方法应用于机械、电机和电子的自动控制系统的研究，形成了工程控制理论的雏形。这一发展过程清楚地表明，工程控制论是控制论应用于工程技术方面而形成的自动控制理论。控制论与工程控制论在学科体系中属两个不同的层次，后者又称为自动控制理论，或简称控制理论。控制论与工程控制论是两个不同的概念，我们必须把它们区别开来。

二　工程控制论的基本内容

工程控制论所包含的内容是随着它的发展而不断丰富的。30多年来，工程控制论的发展已经经历了这样三个阶段，即40—50年代的经典控制理论，60—70年代的现代控制理论和70年代以后的大系统理论。而在每一个阶段中，由于社会生产和科研所提出的任务不同、对象不同，它们所运用的分析方法、数学工具也就不同，基本内容与应用也有区别。

1. 经典控制理论。控制论以前，以伺服机构理论为主要内容的经典控制理论，是以1945年美国波德(Bode)的专著《网络分析和反馈放大器设计》为主要标志的；而在控制论以后，作为工程控制论的经典理论则主要是以1954年钱学森发表的《工程控制论》这一专著为标志。《工程控制论》主要是研究单输入、单输出线性系统，采用的数学工具主要是常微分方程理论、拉氏变换和复变函数论，既有时域法，也有频域法，但以经典频域法为主。对于非线性系统是采用相平面法和描述函数法来进行分析，对于单输入单输出系统的分析方面主要有传递函数法、频率响应法和根轨迹法。

2. 现代控制理论。60 年代前后，主要由于空间技术的进展，如导弹、人造卫星的研制，对自动控制提出了高性能、高精密度的要求，由此，工程控制论的研究对象发生了变化，它必须从理论上与方法上研究和解决多输入、多输出的控制系统、参量随时间变化的时变系统及非线性系统，而且还要解决最优化以及如何适应环境变化的问题。新的问题要求工程控制论有新的概念和方法。于是，1960 年前后，形成了现代控制理论，它的主要标志是卡尔曼提出的状态空间法及能控性与能观测性等新概念，以及同最优控制有关的庞特里亚金的最小值原理和贝尔曼的动态规划法等。这三个方面被称为是现代控制理论的里程碑。这些概念和方法广泛地应用于数学工具，特别是加强了对电子计算机的应用。钱学森在《工程控制论》修订版序言里说："高精度、高可靠性自动调节、自动控制和监测系统，对核能技术的发展极具重要性。"在核电站发展的早期，一般采用常规的机电自动控制技术和仪表。1963 年，在核电站调节、控制与监测工作中首次引用了电子计算机控制，并获得了很大成功。到 60 年代末期，电子计算机监视和控制是核电站技术发展的显著特征。现代电网建设，要求核电站在运行过程中能随电网负荷的变动而自动调整输出功率，只有应用多变量最优控制以及能预测控制变量的前馈控制等现代控制理论，才能实现这个目标。综上所述，现代控制理论的基本内容是以状态空间法为主的线性系统分析，最优控制，自适应、自学习和自组织系统，最优状态估计，随机控制，系统辨识等在工程技术方面的应用，构成了工程控制论的主体。

3. 大系统理论。控制论的对象是系统。所谓系统，是指由相互制约的各个部分组成的具有一定功能的整体。大系统则是指规模巨大的系统，其主要特征是：规模庞大，结构复杂，功能综合，因素众多。所以大系统往往是多输入、多输出、多干扰的多变量系统，有的大系统不仅包含"物"的因素，而且还包含"人"的因素，这类大系统就更为复杂，往往表现为不确定性。一般系统与大系统之间并没有一条严格的划分界限，主要是以上述特征来区分它们。

大系统理论是控制论发展的一个新阶段，它是从 70 年代开始逐渐形成的一个专门领域。由于社会经济、科学技术和军备竞赛的需要，大系统理论已在国际范围内受到了充分的重视。如 1972 年在维也纳成立了国际应用系统分析研究所，这是一个跨国性的研究机构，它研究全球性的大系统问题，如地球资源、人口、生态、能源系统。进入 80 年代以来，大系统的研究和应用发展更为迅速。在理论上，它涉及到大系统的建模和辨识、模型的简化、稳定性、结构和信息研究以及大系统的递阶和分散优化控制等。这些理论的研究成果，已被广泛应用于经济管理、人口预测、交通运输、资源开发、环境保护、空间技术、工业控制等众多领域，并已有许多成功的实例。

大系统的研究不仅与控制论有关，还涉及到许多学科。大系统理论和研究已经成为一个多学科交叉综合的研究领域。此外，大系统还有一个重要特点，就是在控制与管理之间很难找出一条明确的界限。因此，对于大系统的研究，既可以从控制的角度来分析与处理，也可以从管理科学的角度来分析与处理。前者与控制论有关，后者与系统工程、运筹学有关。在此，我们仅从控制理论的角度来简要说明大系统理论的基本内容。

从控制论的角度来分析，一个大系统的控制问题，往往被分解为若干个相互关联的子系统的控制问题。实际上，任意一个大系统的元件、部件与子系统都是要为了实现一定功能而构成一定的控制结构。它们大体上可分为三类：① 多级(递阶)结构。例如一个大企业(公司)的分级管理结构就是一种三级递阶结构。第一级是车间内部运用一些调节装置进行局部控制，称为局部控制级；第二级是工厂运用计算机进行全厂的生产调度，实现过程控制的最优化，称为递阶控制级；第三级是公司的管理计算机，决定各厂控制计算机的最优化指标或策略，进行计划协调与组织管理，称为协调控制级。② 多层控制。是按任务或功能分层，各层之间存在着不同的分工。一般说来，层次越高，任务或功能越复杂，干扰的变化也越快。③ 多段控制。是按被控对象的时序分段，例如，导弹的弹道可以分解为主动段、惯性飞行段和末端制导段。

对于任意一个大系统，一般都能按分级、分层、分段来加以分解。事实上，这就把任一大系统分解为一些不同的子系统，这些子系统之间形成一定的层次关系，表现出大系统内部结构的复杂性。

大系统的模型结构是很复杂的，即使是确定型系统，其数学模型的维数也是很高的，以致于它的求解，即使利用高速的计算机也无能为力，这就是数学模型的维数灾。因此，人们不得不研究大系统模型的降阶问题，或称为模型简化问题，这是大系统研究中出现的新问题。自 60 年代末开始，大系统模型简化的数学方法相继出现，主要有集结法和摄动法。集结法是从经济学中引申而来的一种方法，一般只适用于线性系统。所谓集结，就是加以集中、简化的意思。集结法能将变量很多的大系统模型集结为变量较少的集结模型而保持系统的主要动态特性不变。摄动法是用摄动理论来降低模型的维数。所谓摄动，就是微扰。在一个物理模型中，往往存在一些小参数，这些小参数在整个系统中与一些主要的参数相比，只起微小的作用，但在有些情况下又不能不加考虑，但若把它考虑在内，又会增加方程阶数。因此，微分方程理论中出现了一个分支，即用摄动理论将其阶数降低。摄动理论又分为奇异摄动与非奇异摄动，这种方法不仅适用于线性系统，对于非线性系统也适用，特别对于经济数学模型，奇异摄动法是一种有效的方法。

三　工程控制论的应用简介

工程控制论产生于工程自动控制技术的理论与实践，已被广泛应用于各个领域，从深度与广度上推动了电子计算机技术、核能技术，特别是航天技术和光子技术的发展。

工程控制论在空间技术领域取得辉煌成就，如宇宙飞船的姿态控制和软着陆，使工程控制论的应用达到了目前的一个较高阶段。此外，各种人造地球卫星、行星探测器、运载火箭以及航天飞机、航天遥测、航天遥感、航天遥控以及航天测控信息的远距离传递等等，都是工程控制论在航天技术发展中建立的里程碑。至于工程控制论在工业生产中的应用就为数更多了。□

13 经济控制论

一 经济控制论的形成与发展

控制论已被成功地应用于工程系统和生物系统，那么是否也可以应用于社会经济系统？这是人们共同关心的一个重要问题。

控制论创始人维纳在其著作《控制论》书名的副标题中，明确指出控制论的研究对象是生物领域和技术领域中都存在的控制和通讯过程。在此书的最后一章中，他不很确定地提出一个建议：控制论也许可以用来研究社会现象和思维现象。其后，他在另一部通俗性的著作《人有人的用处》一书中，就是以"控制论与社会"作为副标题的。维纳把社会看成是一种信息控制系统，因此他认为控制论是有必要而且有可能用于社会系统。当然，维纳这里所指的社会是泛指社会的一切领域，不仅仅只是社会经济系统。英国著名控制论专家艾什比在其《控制论导论》一书中，对于控制论研究对象的特点与观点也与维纳的观点基本类似。他认为："控制论也确实可以定义为：它是研究这样一类系统的科学，在这类系统中能量是无关紧要的，而信息及控制却非常重要。换言之，它研究的对象是'不透信息'的。"同时，艾什比强调控制论是研究复杂系统的。对控制论应用于社会研究和管理的必要性与可能性问题，历史上曾有过激烈的争论。经过 30 多年的实践，上述的必要性与可能性已经得到肯定，认识上也基本统一。经济控制论

是控制论应用于社会系统中比较成功和成熟的一个分支，20 多年以来，这一新兴边缘学科吸引了大批不同领域的科学家，使得这一学科无论是理论还是应用方面都有了很大的发展，成为一门与社会经济建设和发展密切相关的重要学科。

对控制论、控制系统的形式化是进行数量化研究的必要前提。控制论对经济控制系统的研究不只是定性的，而且必须是（至少要尽可能地做到）定量。数量化的研究是现代科学技术的突出特点，经济控制论也不例外。有了定量的处理，就有可能进行比较与选择，从而必然出现最优化的要求。最优经济控制是经济控制系统的一项，要求作出最优经济决策。当然，对于一些复杂的、大规模的系统来说，要求实现最优方案是不容易的，在某些情况下只要求满意或比较满意就可以了。

控制论的诞生，提出了一般控制系统的理论和方法。它为研究社会经济系统提供了新思想和新方法，使得经济控制论的产生成为可能。

与所有的边缘学科一样，经济控制论的产生和发展也是有着深刻的历史背景和思想渊源的。马克斯在《资本论》和其他一些著作中，深刻地揭示了社会经济过程中的自动调节与反馈的联系，主张用社会主义计划经济来自觉地调节国民经济的发展，并产生了再生产模式。20 世纪 30 年代，英国经济学家凯恩斯针对当时资本主义世界的周期性危机，主张由资产阶级政府来干预经济管理，并且利用某些经济因素对资本主义经济进行自觉地调节。这些理论和建议，曾为当时的美国总统罗斯福所接受，成为罗斯福新政的一个重要组成部分，企图用来缓解当时的经济危机。从 50 年代开始，英、美等国首先开始将控制论方法应用于社会经济系统的研究。例如，在 50 年代初期，美国加利福尼亚大学教授史密斯（Smith）就运用电子模拟装置来模拟资本主义经济体系，分析研究了其稳定性和对各种干扰的反应。他认为用这种装置解决具体的和一般的经济问题是具有相当大的可能性的。不过，第一本将电子网络理论运用于经济调节与控制问题的专著是英国电子工程学教授图斯丁（Tustin）于 1953 年发表的《经济系统的机制》。他用控制论方框图、信号流程图描述了凯恩斯关于投资、消费、收入之间关系的理论，被认为是经济控制论的创始人之一。在控制论专著中，美国控制论专家彼狄克（Pindvck）是一位较早运用现代控制理论于经济问题研究的开拓者。他在 1971 年发表的论文中，给出了一个包括 28 个状态变量、3 个控制变量（即超额税、政府支出、货币供应量）的美国经济模型。这里显然是运用了现代控制理论，在经济界产生了很大影响。另一位由控制理论转入经济界的杰出代表是柏沙森（Bensoussan）。以他为代表的一批控制论专家认为，控制我们周围经济环境比控制飞船和导弹更有现实意义。经济理论对于控制理论来

说是最有前途的发展方向，到了60年代初，已经有不少经济计量学家和数理经济学家在自己的文献中，运用了控制理论中的一些理论和方法，如最优控制中的三种主要方法：线性规划、动态规划及庞特里亚金的最小(大)值原理，这些理论和方法不仅属于现代控制理论的范围，而且也属于系统工程、运筹学的内容。可以看出，随着现代控制论的研究与发展，一些新的理论与方法，如大系统理论等也被陆续地运用到经济学的研究中来。例如1957年波兰的兰格(Lange)在其专著《经济计量学导论》一书中，已经把线性规划、最优规划这类属于运筹学与系统工程的数学理论运用于经济分析。到了1965年，兰格在其专著《经济控制论导论》一书中，又比较系统地阐述了控制论的基本概念、理论及其在经济领域的应用。此书的特点是，既引用了经济计量学与数理经济学的优秀成果，又分析与研究了控制论应用于马克斯主义政治经济学的问题，而且注意吸收了控制论、系统工程的最新成果，如冯·诺伊曼的可靠性理论等。兰格不仅指出冯·诺伊曼的这一理论可以应用于经济领域，而且用可靠性原理研究了经济系统的效率问题。

自60年代以来，经济计量学和数理经济学不但研究静态和比较静态的最优经济效果及均衡问题，而且发展到研究大规模随机系统的动态最优化，因此，经济学家必然要吸取现代控制理论和大系统理论的理论与方法。这方面的杰出代表是美国普林斯顿经济计量研究中心主任邹致庄，他创立了一个用现代控制理论对动态经济系统进行分析和控制的理论体系。罗马尼亚前总理、经济控制论专家曼内斯库(Manescu)是从另一个角度去理解和发展经济控制论的，他认为经济控制论是一门崭新的边缘学科，是经济学的新理论、新学说。他侧重于宏观经济系统的研究。为了区别于从数量经济学角度发展而来的经济控制论，他主张应从“控制论经济学”的角度来理解经济控制论，而在另一种涵义下的一些研究工作，只能看做是开拓了数量经济学的领域，并没有在基本概念和观念上对经济学进行改造。

“控制论经济学”的上述观点，反映了目前关于经济控制论的两个不同体系，这也正是近年来围绕“经济控制论”涵义诠释的不同观点。

归纳以上分析可以看出，经济控制论是由沿着不同的源头出发的两条相互渗透的途径而形成的。一条途径是，一些控制论学者将控制论的基本概念、理论与数学方法，自觉地运用于经济活动中的自动调节、控制与信息的研究；另一条途径是，一些经济学家，特别是经济计量学家吸收控制论的基本理论，并运用控制理论的数学工具而扩展其研究的领域。简而言之，经济控制论是从控制论和经济学两个不同的源头出发的两条支流汇集而成的，是这两个学科有机结合的产物。此外，经济控制论还与运筹学、信息论等学科有着密切的关系，其形成

是与系统工程、运筹学相互渗透的结果。

经济控制论绝不是简单地搬用了以往经济学中关于调节与控制的思想和数量化的方法，两者是有质的区别的。首先，经济控制论较之以前的一切经济理论，使用了系统的控制论概念、理论和方法，去分析、研究经济活动，把它们看成是一种控制论的经济系统，强调了这类系统的整体性和信息反馈的特点，这是前所未有的。其次，经济控制论不仅是定性地把经济活动看成是一种控制系统，而且运用工程控制论的数学工具与方法对其进行定量的研究，将这两方面有机地结合起来，这也是前所未有的。因此，可以说经济控制论是在质上具有崭新的特点的科学。工程控制论以我国著名科学家、中国科学院院士钱学森1954年在美国发表的专著《工程控制论》为奠基性著作，它已经历了经典控制理论、现代控制理论和大系统理论三个阶段。其中，后二者与系统工程、运筹学相互渗透，在六七十年代出现了经济系统工程的研究，而经济系统工程与经济控制论系统是密切相关的。1965年美国哈佛大学经济系教授杜培尔(Dobell)和控制专业组的何毓琦教授利用控制理论建立了经济模型。1972年由美国国家经济研究局发起，在普林斯顿召开由经济学家和控制理论家参加的首届“随机控制和经济系统”工作会议。1975年在第三届控制论与系统国际会议上，专题报告了经济控制论。至此，社会经济控制理论已成为一个独立的研究学科。

目前，关于经济控制论尚未有统一的严格定义。一般认为它主要是控制论的基本概念、基本理论和方法运用于经济领域(包括经济活动和经济管理)而形成的一门边缘学科。它以各种经济系统的控制问题为自己的研究对象，其基本理论包括经济信息理论、经济反馈理论、经济耦合理论、经济竞争的控制理论以及各种经济放大理论等。它的应用范围极为广泛。随着控制论、经济学及新兴科学理论的发展，经济控制论的内涵必将更为充实和丰富。

二　经济控制论系统

经济控制论系统是经济控制论概念体系的核心，这一概念是由罗马尼亚经济控制论专家曼内斯库提出的，它既是控制论、系统论与经济学相结合的产物，又是经济控制论在“控制经济学”涵义下的独特经济学概念。作为一个系统，经济控制论系统具有一般系统的特征；作为一个控制系统，它又有控制系统的特征，但它是一个特殊的控制系统。因此，用集合论中的术语讲，经济控制论系统概念是上述各种系统概念的“交集”。我们必须从它的特殊性上去认识这一概念的实质。这些特殊性主要表现为：

（1）人的因素。经济控制论系统是一个包括人的因素的系统。如果从控制角度看，人在这个系统中扮演了两个不同性质的角色。一方面，人作为决策者，自觉地直接参与这个系统，人的主观性、目的性都会影响整个系统；另一方面，人又作为决策的执行者，作为一种控制执行因素参与系统，人自觉不自觉地执行决策和调节等活动。人的行为特征及人的心理特征也都会对执行控制产生影响。如果从经济角度看，人又是具有生产和消费双重功能的基本元素。人既是消费者，具有对生产提出目标并体现目标被实现的功能，人同时又是生产者或管理者，具有设计、指挥和执行生产与管理的功能。总之，由于人的因素，使得经济控制系统具有特殊的复杂性，这正是造成一切经济控制论系统具有反馈闭环性质的根本所在。如果我们研究的层次从微观经济逐步扩展到宏观经济，从单级过程上升到多级过程，则所涉及的决策和执行因素将非常复杂。这些因素既可以起促进系统最优运转的作用，又可以起干扰甚至破坏的作用。当然，我们不能因为人的因素的介入而否定认识社会经济控制系统规律的可能性，按照辩证唯物主义的观点，一切经济过程、社会过程、心理过程以及思维过程都是可以认识、可以描述的。

（2）不确定因素。由于存在第一个特点，再加上经济控制论系统一般是巨大的复杂系统，因而随机性、模糊性、偶然性等不确定性因素在这里起着重要的作用。在研究这些系统时，往往要涉及一些非经济因素，例如政策、人口、外国行动、气候等等。这些因素一般具有不确定性。又如在经济关系中，往往具有弹性的特点。例如，当国民收入增加时，某些商品的需求量会相应增加；但当收入减少时，其需求量也还可能有一定的增加。因此可以说经济关系最主要的特征之一是包含着不确定性。随机性、模糊性、偶然性等等不确定因素是经济控制论难以处理的一个重要特点，它使得对这类系统的能控性、能观测性和稳定性的研究更加困难。

（3）大而复杂。从经济控制论系统的结构看，它具有大而复杂的特点。它可以包含许多子系统，每一个子系统由一些有序的相互联系的基本经济元素组成。各个子系统体现了经济上的不同方面、不同层次，各自有相应的子目标，执行不同的子动能。因此，凡是具有某种直接的经济目标，能独立执行相应的经济功能的经济实体，都可以认为是经济控制论系统。一般地，它是一个大系统，所以它具有大系统的各种特殊结构和特征。大系统的控制理论在经济控制论系统中广泛应用。例如，在经济控制论系统中，往往要考虑多目标优化问题，如既要考虑经济费用，又要考虑经济效益，而且还要考虑滞后效应（即经济控制论系统的整体效益往往要经过相当长的一段时间滞后之后才能表现出来）。必须对三者进行科学地对比分析，才能真正达到最优控制或次优控制。因此，在经

济控制论系统中，在最优控制中，要注意经济费用、经济效益和滞后效应的问题，这比一般的最优控制问题又更为复杂一些。

三　建模与控制

在我们对于经济控制论系统进行了分析之后，进而就要考虑如何建立其模型的问题。建立适当的模型并实现最优或次优控制是经济控制论研究的核心问题。

所谓建立经济控制论系统的模型，主要是指数学模型。经济控制论建立其数学模型的目的是要解决最优或次优控制的问题，这一般是通过电子计算机进行仿真，对各种方案进行比较选择后，来解决决策问题的。可以说，经济控制论的基本任务就是解决最优控制与决策问题。但是，由于经济控制论系统是一种有人的因素参与的系统，所以最优控制的问题比较复杂，不仅较难求得，而且有时即使在数据上是最优的，而实际上并不完全合适。因此，一般只求达到次优。

运用自动控制理论、系统工程、运筹学等方法和工具，从整体上研究经济控制论系统的各种特性，在解决能控性、能观察性的前提下，研究其稳定性问题，以便设计相应的施控系统，通过信息反馈对其受控系统进行及时、有效地自动调节和控制，力求通过最优的途径保证上述指标或目标的实现，这就是经济管理问题。给出最优的经济决策，通过最优的经济管理以实现给出的经济指标，就是最优的经济控制。这是经济控制论的主要任务。

应该指出，经济控制论不仅要求达到最优控制的目的，而且要求在整个的经济过程中能进行有效的自动调节与控制，以保持系统的稳定性，通过最佳途径达到最优控制的目标。在这方面，它比经济计量学是有所发展的。比如，经济预测早已成为经济计量学的重要内容之一，与之相比，经济控制论则不仅能更有效地进行经济预测，而且能有效地进行自觉地调节与控制，以实现最优控制。在一定的意义上可以认为，只有既能对最优目标进行预测，又能通过自觉地自动调节与控制，通过最佳途径来实现预测，才算全面地实现了最优控制。经济控制论之所以能够做到这一点，从理论上说，很重要的原因就在于运用了信息反馈的原理，从而使它既与以前某些经济理论关于控制关系的认识相区别，又与经济计量学等相区别，而显示其优越性，保持其独立性。应该重视的是，信息处理问题是控制的必要条件，因此，为了实现最优的经济控制，必须建立起相应完善而合理的经济信息系统。

曼内斯库认为，通常一个模型的制作过程表现为一组三元体$\{A,S,M\}$，A

是原系统，主体 S 在模型 M 的帮助下对 A 进行研究。他将模型的制作过程分为三个不同的阶段：第一阶段向模型 M 传递所有关于被分析系统 A 的已知信息；第二阶段，在已知的科学工具帮助下对模型 M 进行研究，获得关于系统性能的新信息；第三阶段，通过研究模型和在实践中使用模型获得的新信息输入模型系统 A。经济模型的建立，通常是从建立定性模型入手，在定性模型的基础上，采用各种数学方法（如统计方法等）将定性模型中的部分性质定量化，并利用计量经济学方法和最优控制理论确立各性质和变量之间的定量关系，从而初步建立定量的经济模型。模型的建立并不是一次完成的，有时还需要经过多次修正和检验后才能最后确定。

经济系统的控制方法。经济控制系统是一类特殊的控制系统，因此关于一般控制系统适用的原理和方法，如最优控制、随机控制、模糊控制及大系统控制的理论和方法，原则上都可以应用于经济控制论系统的控制。但是由于经济控制论系统的特殊性，使得经济系统的控制问题也具有某些特殊之处。

首先在经济系统最优控制问题的提法上具有一些特点。例如，对于宏观经济系统而言，控制的目的在于实现高速度发展的社会经济，将国民经济的发展控制在最优、平衡、按比例增长的轨道上。所谓“最优”，是在经济效益的涵义下实现系统的优化控制；所谓“平衡”，是指控制的稳定性和适应性；所谓“按比例”，是指控制的统筹协调作用。对于微观经济系统而言，控制的目标通常是最大收益和最小耗费。其次，由于经济系统模型既有定性模型，又有定量模型；既有确定性，又有随机性和模糊性。因此，确定控制策略的方法具有多样性，定量和定性方法相互结合；在研究方法上，具有“软科学”方法和“硬科学”方法结合的特点，这种特点使得确定经济控制系统的最优控制策略更为困难和复杂。另外，经济控制系统一般是大系统，大系统中的分解—协调原理和分散控制方法被广泛应用。经济控制系统中的滞后效应也往往比较明显，控制作用的效果往往经过较长的时间才能表现出来。

经济控制系统一般都是非平衡系统。从空间看，产业之间、部门之间、地区之间的经济发展是不平衡的；从时间看，经济发展的速度是非均匀的，有规则的经济波动和无规则的随机变化相互影响，造成经济发展过程中的涨落现象。这种非平衡经济控制系统的控制策略更为复杂，这也是经济控制论中亟待研究的新课题。□

14 独具特色的黑箱方法*

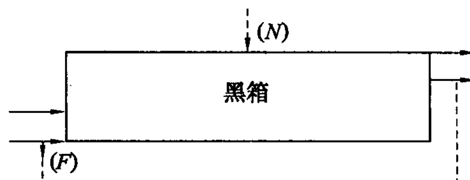
浩瀚的宇宙有许多发人深思的复杂事物，如人的大脑是如何进行思维活动的，人体上共有多少个穴位，每个穴位有哪些物理的、化学的、生理的特性，各有什么功能，人是怎样致癌的，人能否控制癌等等。这些问题犹如一个个既不透明又密封着的箱子。在研究它们的时候，由于条件的限制，又不容许或不可能把它们打开，因此不能直接让人们洞察它们的内部结构。例如，倘若为了研究人脑而让一个脑外科医生对人的大脑实施手术剖析，就如今的医学技术水平而论，不仅意味着可能会失去这个人最宝贵的生命，而且还因为只有当人脑还活着，并且还在工作或活动的时候，才能表现出它的功能来，死了的人脑就失去了一切功能，也就失去了机制，因此也会一无所得。像人脑这样结构复杂、机理奥妙的复杂事物，在 20 世纪 40 年代，美国控制理论学家维纳把它看成是一个黑箱(blackbox)或黑色系统，并选用了所谓黑箱方法来研究人脑神经细胞系统的功能和行为，获得了丰硕的成果。科学工作者触类旁通，对此成果加以发展，把那些人们需要认识的事物，但由于当时条件的限制，其内部结构情况不能或不容易直接观测到，就像一个既不透明又密封着的箱子，都叫做黑箱。黑箱的例子俯拾皆是，如宇宙科学、生命现象、生态系统、原子结构、密封着的仪器等等。

与黑箱或黑色系统相对的，是所谓白箱或白色系统。凡内部特性确知或能知的事物或系统，叫做白箱或白色系

*本文原载于《自动化博览》1989 年第 2 期。

统。1983年，我国邓聚龙教授又提出了“灰色系统”(不妨称之为“灰箱”)。所谓灰色系统，是指既含有已知的信息，又含有未知的或非确知的信息的系统。如一个工厂是一个系统，工厂的人员、设备、技术条件是已知的。根据这些已知的数据可测算出它的产量、产值，这样的系统称为白色系统。宇宙间，银河以外的某星球，虽然知道它存在，但不知其质量是多少，体积多大，距地球多远，是否发光，是恒星还是行星，这就是黑色系统。人体是一个系统，人体的外形参数，如身高、体重、年龄等都是已知的；某些内部参数，如体温、血压、脉搏等都是能知的，这些都是白色参数。但还有更多的参数是不知的，比如人体上究竟有多少穴位，每一个穴位究竟有哪些物理的、化学的、生物的性能，这些穴位与人体的哪些部分有关，它都起什么作用等，都是未知的或者是不确知的，是黑色参数。系统中既有白色参数又有黑色参数，就称为灰色系统。一些抽象系统，如生物系统、农业系统、社会经济系统都缺乏确切的物理原型，它们的系统信息和系统参数不完全，部分清楚，部分不清楚，都属灰色系统。

黑箱方法又称为黑色理论，首先是把研究对象看成是一个整体，看做是黑箱，建立黑箱和外界环境的输入输出关系。所谓输入关系是指在黑箱方法中，把受到的周围环境的影响看做是外界向黑箱输入信息，而把研究对象对周围环境的反影响看成是黑箱向外界输出的信息。在黑箱中能反映输入、输出真实情况的，只能是那些不断涌现的信息流，这种关系可以用框图表示。下图中，(N)代表在黑箱试验中有时遇到的环境干扰影响，(F)代表输出反馈影响。



当输入输出关系确定后，再通过观测试验，利用研究对象的特征性质和对研究对象有影响的各种因素数据建立数学表达式。但这只是一个近似的数学表达式，还要利用这个近似的数学表达式进行预测，把预测的结果跟再次试验所得的数据进行多次的反复调整、修正。经过有限次数的信息输入、输出的循环过程，就可以得到关于研究对象的正确的数学表达式。有了正确的数学表达式，就可以对黑箱的内部构造机理、功能和特性进行定性、定量和静态、动态的分析和评价，并作出推理和假设。这样去认识事物，使黑箱变成白箱，进而达到控制的目的。因此，所谓黑箱方法不妨说是一种特殊的数学方法，即所谓构造方法。

由以上所述可见，黑箱方法并不需要打开黑箱，而是通过外部观察和试验，

就可以了解事物内部情况的变化。它是从事物的整体功能出发，不考虑事物的局部细节。

黑箱方法最初只局限于控制理论的研究，近年来，已推广应用于难度较大的工程技术、生物、医学和社会经济等不同的学术领域内。因此，黑箱方法有较广泛的应用价值，解决了一批一向被认为难以解决的问题。

黑箱方法虽然在 20 世纪 40 年代才出现，但它在人类历史上可谓源远流长，如众所周知的我国中医理论用的就是黑箱方法，它距今已有两千多年的历史。中医为病人诊病，不是将人体分裂开来，而是将其看成是一个整体，即从整体的功能出发，来研究人体生理和病理的基本规律。在诊病的时候，是通过望、闻、问、切这四种方法，得到人体这个黑箱输出的信息，去诊断人体功能和疾病状况。历史上很多著名的研究成果，实际上也是由黑箱方法获得的。例如在研究高级神经活动，即大脑的机理时，俄罗斯生物学家巴甫洛夫采用的就是黑箱方法，如把铃声看成是对狗脑输入的信息，把狗脑受铃声刺激所分泌出来的唾液看成是狗脑输出的信息，通过对狗脑输入和输出信息的分析，研究脑的生理、心理活动的机制，创立了著名的“条件反射”学说。即使在日常生活或工作中，我们也时常不自觉地应用着黑箱方法，例如列车的检车人员用锤子敲打车辆部件（输入），通过倾听发出的声音（输出），来判断火车车体是否有故障。

英国控制论学者艾什比曾说过：“所有的事物，实际上都是黑箱。并且，我们从孩提时代起直到老态龙钟，一辈子都在跟黑箱打交道。”黑箱和白箱不是什么绝对不变的东西，在辩证唯物主义者看来，在一定条件下，黑箱是可以转化为白箱的。

黑箱方法的应用范围很广，有着独特的优点，但也存在着局限性，就是强调了外部观测和整体功能，而忽视了事物内部的精细结构和局部细节。所以只有把黑箱方法同其他的科学方法结合起来，取长补短，才能对客观事物有全面的深刻的认识。

黑箱是模糊的，不能给人以清晰的信息。自从 60 年代中叶，美国控制学家扎德发表了题为《模糊集合》的开创性论文，80 年代后，经过许多人的努力开拓，已初步形成了一个新的数学分支，即模糊数学。近年来，已有人把模糊数学的理论和方法，试图用以解决黑箱问题。果真成功，那就为把黑箱变为白箱开辟了一个新的途径。□

15 黑箱、灰箱和白箱方法*

随着现代科学技术的发展，越来越多的事物成为人们研究的对象，研究的方法也日新月异，如为了提出和解决被控系统的建模问题，控制学家维纳和艾什比提出了黑箱理论和黑箱方法，它与灰箱、白箱方法一起发展成控制理论的一个重要分支，即系统辨识的理论基础。

一个系统，不知它的内部结构如何，像一个既不透明又密封着的箱子，故称黑箱或黑色系统。在研究它的时候，由于条件的限制，不能把它打开，直接洞察它的内部结构，只能靠对它的输入和输出信息来研究黑箱的功能、特性，这就是所谓黑箱方法。1956 年，艾什比在其所著《控制论导论》一书中对黑箱理论和黑箱方法作了系统的阐述："黑箱是在电机工程中出现的，给电机工程师一个密封箱，上面有些输入接头，可以随意通上多少电压、电击或别的干扰；此外，还有些输出接头，可以借它做它所能做的观察。"这个密封箱就是一个黑箱。

黑箱问题是广泛存在的。例如人脑就是一个著名的黑箱。虽然我们知道它具有记忆、联想、学习等功能，但实现这些功能的生理、生化、形态的内部机制，以及脑的内部精细结构如何，直到今天，人们知道的还不多。维纳就是把人脑看成一个黑箱，选用了黑箱方法来研究人脑神经细胞系统，获得了丰硕的成果。

艾什比认为，控制论的贡献不在于把这样一类我们对其一无所知的系统称之为黑箱，而在于它提供了一种认识

---

* 本文原载于《知识工程》1992 年第 2 期。

这种黑箱的方法，即不需要打开黑箱，而是通过外部观察和实验就可以了解系统内部情况的变化；同时，它是从系统的整体功能出发，不考虑系统的局部细节。

近年来，黑箱概念和黑箱方法已被广泛应用于难度较大的工程技术、生物科学、社会和经济等领域。例如，由于人的心理系统十分复杂，打开这个“黑箱”十分困难，所以在心理研究、智力测验上，研究者基本上是采用黑箱方法，即在言语、动作方面输入信息，观察人的输出，而不是通过解剖脑和神经系统来研究人的心理状态或智力状况。

由此可见，黑箱方法具有明显的优点。由于科学的研究总是在探索那些未知的领域，所以就是不断地打开各种黑箱。在探索那些不需要打开或无法打开的未知系统时，黑箱方法不失为一种有效的方法。但是，黑箱方法也有其局限性。第一，对功能上的“同构”现象，它可以作出许多种假设，很难确定哪种假设是正确的。其次，黑箱方法强调了外部观测和整体功能，而忽视了事物内部的精细结构和局部细节。所以应当把黑箱方法同其他科学方法（例如功能方法与要素结构方法）结合起来使用，取长补短，才能对客观事物有全面的、深刻的认识。

黑箱方法以我们对所研究的对象“一无所知”为出发点。人类对客观事物的认识是从不知到知，从知之甚少到知之较多，是一个不断深化、不断积累的过程。人类历史的长河表明，任何一个人都是在直接接受了前人的经验和知识的前提下，研究新的课题和对象的。例如，从控制论的角度说，我们面对的是对其有了部分了解，但又不完全了解的黑箱系统。也就是说，它不是一个完全的黑箱。艾什比把一部分了解的黑箱或黑色系统称为“部分可察黑箱”。我国华中理工大学教授邓聚龙于 1983 年提出的“灰色系统”理论，其实就是这种“部分可察黑箱”。为与白箱、黑箱相对称，我们在这里将其称之为“灰箱”或“灰色系统”。也就是说，系统中既含有已知的信息，同时又含有未知的信息，则称此系统为灰箱或灰色系统。在灰箱中，我们具有关于实现输入输出关系的结构和功能的部分信息，例如法院办案，就案件来说是一个系统，审判人员对作案者的作案信息有一部分是确知的，又有一部分是不知的或非确知的，审判就是对案件逐步加深了解并全面认识的过程。只有对案情分析、立案，取得人证、物证，把案件全部弄清楚，审判人员才能依法作出正确的判断，即决策，才能真相大白，所以法院办案就是一个典型的由“灰”变“白”的过程。

灰箱也是普遍存在的，如生物系统、农业系统、社会系统、经济系统等，都无确切的物理原型，它们各自的系统信息部分清楚，部分不清楚，所以都是灰色系统或灰箱。在一定意义上说，进一步完善灰色系统的数学理论，有可能发展成

灰色数学，并将有广泛的应用前景。因此对灰箱的研究，不但具有理论价值，而且也具有重要的实际应用价值。

解决灰箱问题，要应用相应的灰箱方法。灰色系统的一个基本观点认为，客观对象尽管复杂，数据没有明显的规律，但它必定是有序的，有内在功能的。因此可将原始信息数列通过一定的数学方法进行处理，将其转化为微分方程描述的客观规律。灰色系统理论对数据的处理通常是用累加或累减生成方法，使无序数列转化为有序数据数列，并使数据序列适宜于用微分方程建模。这种使系统信息由不知到知之，或由不确知到确知的过程，就是系统由“灰”变“白”的过程。它解决了一向被认为不能解决的连续微分方程建模等问题，为自然科学与社会科学的沟通，为硬科学与软科学的量化研究和模型建立，特别是为未来学的研究提供了新的理论和方法，这是在科学上的重大突破。因此，在短短的几年里，经过创始人和他的同伴们的共同探索，灰色系统已初步形成了一门横断面较大、渗透较强的新的边缘学科，并被广泛应用于工农业生产、社会、经济、生态等领域的研究中。

灰色系统理论目前主要包括灰色系统分析、灰色系统建模、灰色预测与决策、灰色控制等分支。下面对灰色预测与决策作简要介绍。

一些经常重复出现的事情，其产生的背景、特点，有关的各种因素和未来发展的趋势，全都清楚明白，决策者只要对已知的数据进行既定的计算，或编成程序通过计算机求解，就可作出正确的预测与决策，这叫白色预测和决策。但现实世界中，由于各种条件的限制而无法认识某事物的全貌和相关因素。对于这类“因素不清、结构不明、数据不全、难以模型化”的事物进行观测和决策，就称为灰色预测和决策。比如有人从武汉到重庆，可以乘飞机、火车或轮船，这里有三个可供选择的方案，哪一个最佳呢？这要看效果，而效果又与目标有关，不同的目标对效果的评价是不同的。设要考虑三个目标：省钱、省时和方便，从省钱来看，乘轮船最好，但省时方面最差；从省时来看，乘飞机最好，但省钱最差；从方便来看，火车一天通常有几班，但省钱与省时介于飞机与轮船之间。这样的决策便是一个灰色决策。灰色决策包括灰色局势决策、灰色线性规划、灰色层次决策等，都有广泛的应用。

与灰箱方法的应用相联系，人们又提出了白箱方法。维纳在 1962 年再版其名著《控制论》中明确提出：“我们把某些具有已知结构的物体称为白箱，用它们表示所求展开式的各项。”这里所谓“展开式的各项”，维纳认为就是“白箱网络”。解决白箱问题，相应地要用白箱方法。所谓白箱方法，据维纳解释就是这样的一种方法：当我们通过诸如黑箱方法、灰箱方法等认识了系统的内部结构时，就可以把这种结构关系按一定的关系式表达出来，这就是白箱网络。其实，

制定白箱网络并不是白箱方法的全部目的，更重要的是通过这种白箱网络对系统进行再认识，或者利用这种白箱网络去控制系统以后的过程或预测系统的行为。例如，我们使用电子计算机，其运算程序都是事先设计好的，这种程序就是以上所说的白箱网络。因为在设计程序时，我们对于程序输入计算机后，它的运算步骤、运算方式都是知道的。由此我们可以看到，白箱方法反映出比黑箱方法、灰箱方法更高一级的认识水平。一般地讲，完全的黑箱，我们最初对它是不可能有任何预测性的，但到了灰箱，预测性则随之增加，而在白箱理论中，预测性则相对地达到了最好的程度。

总的来讲，黑箱方法、灰箱方法、白箱方法现在已经成为控制论中的一个重要分支，即系统辨识的理论基础。系统辨识，顾名思义，就是辨识、测量与认识系统。精确点讲，它要解决如何正确建立动态系统数学模型的问题，而采用的主要方法是测定输入、输出数据，然后通过这些数据确定系统的结构与参数，从而求得定量描述系统的数学模型。而系统的输入、输出数据中包含了系统结构的信息，我们知道，这正是黑箱方法的观点。但是，由于结构与功能的错综复杂，人们很自然地提出了系统结构的可辨识性问题，即系统辨识问题，它被广泛应用于工业、国防尖端科学技术及经济学、生物学、医学等众多的领域。□

16 系统辨识浅谈*

一　系统辨识的基本理论

20 世纪 50 年代发展起来的经典控制理论被成功地应用于单输入单输出、定常系统的分析与设计。60 年代初，由于航天航空技术的需要，控制论有了进一步的发展，从经典理论发展到了以状态空间方法、卡尔曼的滤波理论、贝尔曼的动态规划方法及庞特里亚金的最小值理论为主要内容的现代控制理论。它具有更广泛的适应性，可用于多输入多输出、定常或时变系统的分析与设计。在系统的性能指标上，现代控制理论提出了“最优”的概念，最优控制理论取得了长足的进展。所有这些理论都建立在受控系统的数学模型上。但在很多情况下，建立受控系统的数学模型并不是轻而易举的，尤其是具体的物理对象或工程系统，它们的机理十分复杂，而且还含有各种噪声，这就使建模更加困难。因此系统辨识(identification)，即用实验取得的输入输出数据建模就应运而生。由于近十几年来，计算机的发展与普及为系统辨识的进一步发展铺平了道路，使系统辨识成为自动控制理论的一个重要分支，受到控制理论工作者们的热切关注。系统辨识是现代控制理论研究的基本课题之一。

什么是系统辨识，目前还没有公认的定义，也有人译

* 本文原载于《自动化博览》1990 年第 5 期。

为系统测辨或系统识别。顾名思义，它就是要辨别、测量与认识系统。1962 年，美国控制论专家扎德在其《从线路理论到系统理论》的论著中，曾给系统辨识下过定义："系统辨识就是在输入和输出数据的基础上，从一组给定的模型类中，确定一个与所测系统等价的模型。"从理论上讲，这个定义无懈可击，它明确了系统辨识的三个基本要素：① 输入输出数据。把输入输出所观测到的信息统称为输入输出数据。② 模型类。即按辨识的目的要求，在先验知识的基础上确定被辨识系统模型的类型。③ 等价性准则。等价的概念指的是，如果两个系统在所有可能的输入下具有完全相同的输出数据，则称这两个系统等价。对象和模型的等价性准则是通过引入评价函数来定义的。由于评价函数是用来衡量对象和模型之间的相似性的，所以人们希望它具有物理意义，而且使用起来简单。

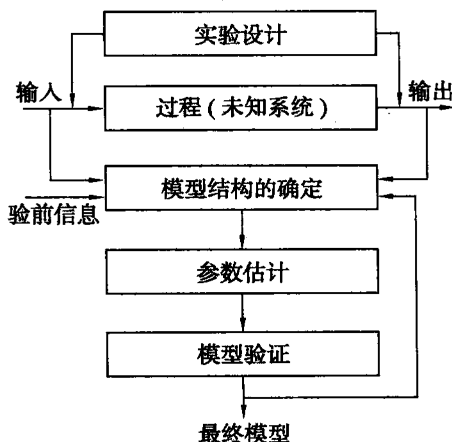
在上述三个要素中，数据是辨识的基础，准则是辨识的优化目标，模型类是寻找模型的范围。但事实上，按照扎德的定义寻找一个与已给系统完全等价的甚至近似的模型并非易事。幸好，从实用的观点出发，对模型的要求并非如此苛刻。为此，对系统辨识又有一些比较实用的定义。比如 1978 年，Ljung 给辨识下的定义："辨识有三个要素——数据、模型类和准则。辨识就是按照一个准则在一组模型类中选择一个与数据拟合最好的模型。"总而言之，辨识的实质就是从一组模型类中选择一个在某种准则意义下最能代表该系统特性的数学模型，这种建模的方法就称为系统辨识，它是在不知道系统内部的结构是什么，只利用实验的输入输出数据来建立系统的数学模型的一种方法。所以如果对系统一无所知，则系统辨识就是黑箱建模方法。辨识黑箱是比较麻烦的事，幸好人们总是可以对系统做一些调查，对它的情况有先验部分的了解，黑箱问题便退化为灰箱问题。辨识灰箱的数学模型相对地较为容易。除了系统辨识这种建模方法以外，还可以利用分析方法建模。分析法是把受控对象(系统)分解成若干子系统或环节，分别根据某些已知的物理、化学、生物和其他基础学科的定律、公式等列出数学关系式，考虑到各个环节之间的相互关系，从而推导出受控系统的数学模型，其被称为机理模型。对复杂受控系统，其机理模型事实上很难建立，在这种情况下可用系统辨识方法，即通过实验数据建模。但辨识建模所得到的模型只是反映输入输出的特性，对系统内部的信息反映不出来，而机理模型则可以弥补其不足。所以，在条件许可下，应该同时把系统辨识方法和分析方法有机地结合起来建模，以便取长补短，互相补充。

由以上所述可以看出，系统辨识的任务就是将通过实验得到的输入输出数据，按一定的目标函数，从一类数学模型中确定出一个数学模型的结构。模型的结构确定后，由于模型还往往包含若干待估计的参数，所以还有相应的参数确定问题，这就是所谓参数估计。等到把参数都估计出来了，模型才算最终建

立起来。参数估计的方法很多，如有以脉冲响应为基础的脉冲响应法、相关函数法和局部辨识法，有以最小二乘法为基础的加权最小二乘法、递推最小二乘法、广义最小二乘法和辅助变量法等，有以似然函数为基础的极大似然法，另外还有时间序列的建模和随机逼近法等。要针对不同类型的模型，选用不同的参数估计方法。

参数还可分结构参数与模型参数。所谓结构参数，对于单输入单输出系统来说，通常指的是它的阶数；对于多输入多输出系统来说，不同的模型形式有着不同的结构参数，即使采用状态空间模型，不同的标准型的结构参数也是不同的。例如，当采用可观测标准型时，其结构参数是可观测性指标。所谓模型参数，是指不依赖输入输出和状态变量的那些参数，例如传递函数、微分方程、差分方程的诸系数和状态方程中的各系数矩阵。

最后，当辨识提出了对象的数学模型后，它究竟在多大程度上真实地反映了对象的动态特性，还必须通过检验，这就是所谓模型验证。比较常用的验证方法有"白色度"检验法。当然也可以根据特定的目的要求，通过实验来加以检验。例如对模型与对象施加相同的输入信号，比较它们的输出，如果偏差很大，则修正模型。如此反复循环，直到满足一定的精度为止。这样，系统辨识的内容包括：系统辨识的实验设计、模型结构确定、参数估计、模型验证等。系统辨识的一般步骤可用右图表示。



二　系统辨识的基本方法

1. 以脉冲响应为基础的辨识方法

脉冲响应法是一种比较实用且具有一定范围的辨识方法，它利用线性、定常系统的不包含噪声的输入输出信息，通过脉冲响应来辨识系统的数学模型，既是一种非参数型（脉冲响应）的辨识方法，又是一种通过脉冲响应而得到参数模型（传递函数）的辨识方法。这种辨识方法的优点是简便易行，主要缺陷是它仅适用于确定型线性定常系统，在有随机噪声的情况下，会带来辨识误差而影

响到结果的精度。

2. 以最小二乘法为基础的辨识方法

18 世纪，高斯为了解决观测得到的行星轨道的数据推算行星轨道的参数，提出了最小二乘理论，人们称之为一般的最小二乘法。从此之后，最小二乘理论便成为处理观测所得数据的有力工具。在系统辨识领域内，最小二乘法成了参数估计的最基本方法。考虑到有的观测数据影响大，有的影响小，因此又引入了加权最小二乘法。到了 20 世纪 60 年代，计算机技术的发展，推动了递推算法的广泛应用。递推算法不仅实施方便，更重要的是它适用于在线辨识和实时辨识；既适用于单变量线性系统，也适用于多变量线性系统。由于最小二乘法简单实用，它又是构成其他类型辨识方法的基础，因此它是一种基本的、重要的辨识方法，例如，在最小二乘法的基础上，作了一些改进而产生了辅助变量法、广义最小二乘法、时变参数系统的辨识方法等。□

17 大系统理论及其优化方法*

大系统理论是控制论发展的一个新阶段，是现代控制论发展的一个新的重要领域。

1972 年，在维也纳成立了“国际应用系统研究所”，这是一个跨国性的研究机构，它研究与全球有关的一些大系统问题，如地球资源问题、人口问题等。1976 年，国际自动控制联合会召开了大系统理论及应用的专题讨论会，这表明大系统理论已在国际范围内受到了充分的重视，并迅速发展起来。

那么怎样的系统才是大系统呢？大系统有哪些特征？这是大系统理论研究首先要回答的问题。一般来说，大系统具有如下一些特征：

（1）系统模型的维数高、规模庞大。主要是指元件、部件多，包含着许多彼此关联的子系统；方程的个数、约束条件、状态变量有数百或成千上万个。所以大系统一般是多变量系统。

（2）系统结构复杂。主要是指系统包含许多相互关联甚至是相互矛盾的子系统，整个系统的特性既体现在各个子系统单独的特性上，同时也体现在它们之间相互关联的特性上。

（3）系统功能综合。表现为大系统往往具有多个目标，其中的某些目标之间可能还是相互克制或矛盾的，需要多目标决策和优化的方法。

（4）系统因素众多。主要是指大系统所处的内外环境

---

* 本文原载于《自动化博览》1991 年第 4 期。

十分复杂，受到的干扰和影响多，所以大系统往往是多输入、多输出、多干扰的多变量系统。在有的大系统中，不仅包含“物”的因素，而且包含“人”的因素。有的其时空分布特殊，例如一个多级的生产计划系统；有的系统可能沿地理分布很广，例如河流污染的控制系统。

(5) 随机性。如系统的资源或运行要求常带有随机性，系统部件故障的发生也有随机性等等。

根据以上特征可知，对于大系统，我们可以理解为变量数目多、结构复杂、功能综合的多输入、多输出、多干扰的多变量大型系统。大系统理论不仅与现代控制论有关，而且与运筹学和许多其他学科也密切相关。大系统理论已成为一门横跨许多专业领域的边缘性多学科交叉科学。

从控制理论角度分析，任意一个大系统的元件、部件与子系统都要为实现一定的控制功能而构成一定的控制结构。对于大系统来说，无论因素多么众多，结构多么复杂，都必须对它进行定性与定量的研究。从数学上讲，大系统研究的主要是大系统的优化，这就要首先建立它的数学模型，用数学表达式来刻画系统的各种性能。但由于大系统的因素众多和结构复杂，使得数学模型难以建立，即使建立的是确定型的数学模型，其方程和阶数一般也是较高的，可能达到几十、几百甚至上千阶的高度，这就是大系统数学模型的维数灾难。同时，出现的方程个数也可能是大量的，用一般的最优化方法和计算技术，在目前的计算设备下解决大系统最优化问题会遇到很大困难。为了克服求解上的困难，人们对某些大系统的最优化，从实践中总结出了一些行之有效的方法，大体上可分为三类：

1. 直接法。即利用大型优化问题的特殊结构，改进现有的优化方法用于求解一类特殊的问题。如线性规划中的单纯形法，非线性规划中的共轭梯度法、拟牛顿法等。在大型线性规划方面，其约束系数中的非零元素通常组成稀疏矩阵(稀疏矩阵是指它的非零元素所占的比例很小，一般不超过所有元素的0.5%，而且按照一定的规律排列，即其非零元素往往组成阶梯形、对角块等各种特殊的结构)。直接法就是利用结构的特殊性，以达到减少计算时间、减少存储的目的。具体地说，大型线性规划的直接解法是指单纯形法在特殊问题中的推广，其处理问题的关键是利用紧缩基逆方法，对基逆矩阵进行修正。紧缩基逆方法是单纯形法的变种，它的主要优点是计算量少，要求较少的高速存储器，因此被广泛应用于解大型的线性规划问题。

2. 分解—协调方法。大系统是由许多彼此关联的子系统所组成，每一个子系统的阶数比整个系统的阶数要低得多。子系统的求解不仅比大系统容易，且仅需一台价格便宜的小型计算机即可解决问题。因此一类重要的大系统优化

方法就是分解—协调方法。这种方法的基本思想是：首先把一个大系统分解为一些孤立的和较小的子系统，分散求解子系统的最优化，以代替集中或全局最优化的方法。但局部最优化并不等价于全局最优化，所以，必须在子系统最优化的基础上，反复协调局部的最优解以得到大系统的全局最优解。因此对于求解大系统最优化问题，既不能和小系统的所谓"一揽子"求解方法一样，也不能只是单纯的分散求解，而是既分散又协调，如此逐次迭代的分解—协调，最后便收敛到全局的最优解。分解—协调就是求解大系统的基本思想和方法。实际上，对于多目标的大系统来说，难以做到真正的最优化（或达到最优控制），一般只要做到合理兼顾、保证重点就算满意了，这叫做大系统的"次优化"或"准优化"。

3. 大型线性规划的多项式算法——卡玛卡算法。1947 年，美国年轻的数学家丹茨格发明了求解线性规划的单纯形法，实践证明，只要变量个数不超过 20 000，使用起来还是有效的。因此，单纯形法被制成标准软件，广泛应用于各个领域。但自 70 年代以来，它已被证明并不是一种好算法。通常评价一个算法好坏的标准是，它在得到最终答案之前，需要运算的次数多不多，或者说运算得快不快。一个算法，如果运算次数是按指数函数增长，则称该算法是指数时间的，这样的算法便被认为是慢的算法；如果运算次数是按多项式函数增长，则称该算法是多项式时间的，这样的算法便被认为是快的算法。单纯形法是不是多项式时间算法？回答是否定的。1972 年，克利和明蒂以及以后其他一些人都已证明单纯形法及其变种全是指数时间的，所以从计算复杂性角度来看，单纯形法并不是一种好算法。计算机必须用一种最佳算法去检验尽可能少的解，从中找到最优解。1984 年，印度青年数学家卡玛卡发明了一种求解线性规划的新的多项式时间算法。1988 年在东京举行的国际数学规划讨论会上，卡玛卡算法的有效性得到了公认。对同样的问题，卡玛卡算法不仅比单纯形法用时更少，并且不会出现用单纯形法求解时有可能随着问题规模的增大而使计算量按指数规律增加的情况。卡玛卡算法是求解线性规划的重大突破，为求解大型的复杂线性规划问题提供了实际可行的新工具。□

人工智能：最精彩的人类智慧产品

1 人工智能及其新发展*

所谓人工智能(artificial intelligence，简称 AI)，是指把人的某些智能赋予机器，让机器模拟和代替，去做原来只有人才能做的智能工作。人工智能是相对于人的天然智能而讲的。由于是用机器来实现人的某种智能行为，因此人工智能也称为机器智能。

人工智能是探索和模拟人的感觉和思维过程的规律，并进而设计出类似于人的某些智能的自动机的科学。由于电子计算机的发展已具有了逻辑推理功能和存储功能的能力，所以当前这种探索、模拟人的智能活动的主要工具是电子计算机。1977 年，美国人工智能专家温斯顿(Winston)在其所著《人工智能》里说："人工智能就是研究如何使计算机去做原来只有人才能做的具有智能的工作，人工智能的中心目标就是使计算机更有用以及探讨构成智能机器的原理。"事实上，人工智能采用的是计算机软件的观点和方法，它已经成为计算机科学的一个分支学科，主要目的是用现有的计算机去探讨智能的原理，使现有的计算机更为聪明，更为有用，进而研制出高级的智能机。所以人工智能不仅是自动化技术的前沿，也是计算机发展的方向，是控制论的尖端问题。

然而，究竟什么是智能，给它下一个确切的定义是困难的。概括地说，智能通常称为智慧，它"是指人认识客观事物并运用知识解决实际问题的能力，集中表现在反映客观事物深刻、正确、完全的程序上和应用知识解决实际问

* 本文原载于《自动化博览》1991 年第 5 期。

题的速度和质量上”。人类的天然智能伴随着人类的活动到处存在，随着时代的发展，人的智能也不断发展，因而智能的概念也不是一成不变的，它是一个历史的范畴，具有明显的时代特征。人类的许多脑力劳动，如解题、编写计算机程序、证明定理，甚至驾驶汽车和骑自行车等，都需要不同程度的智能。如果计算机或其他机器能够执行这种任务，就可以认为这类计算机或其他机器具有某种程度的人工智能。简单地说，由计算机来表示和执行人类的智能，就是人工智能。或者更确切地说，人工智能就是使计算机进行某些与人的智能有关的功能（如图像识别、问题求解等）的能力。具有这种功能的机器，就称为智能机。在过去的二三十年中，已经建立起了一些具有人工智能的计算机系统，特别是有的计算机系统已能够诊断疾病、求解符号形式的微分方程、分析电子电路等。我们可以说，这样一些系统具备了某种程度的人工智能。

人工智能同现代控制理论、逻辑学、心理学、算法分析、计算机科学、仿生学、信息论、运筹学等学科密切相关，并相互渗透，所以人工智能是新兴的且在发展中的边缘性综合科学，它的研究范围十分广泛。尽管到目前为止，还没有一个明确的界线，但就其本质来说，凡使计算机产生人类智能的基本原理的研究以及使计算机实现人类的某些智能的研究范围，像自然语言处理、图像和物体的识别（或称模式识别）、知识表达、机器定理证明、机器人学、专家系统等，都是人工智能的不同分支，而 LISP 语言则是人工智能最常用的语言。今天，人工智能与空间技术、能源利用一起已成为现代科学技术领域的三大重要标志。

机器定理证明。证明定理是人类特有的智能行为，是逻辑演绎的过程。机器定理证明的过程就是把人证明定理的过程通过一套体系符号加以形式化，变成一系列能在计算机上实现的符号演算过程，也就是把具有智能特点的推理演绎过程机械化。

许多数学分支中都有一些欲证而不能的猜想，对数学中一个猜想的定理寻找一个证明，这确实可以称得上是一项智能的任务。1956 年，美国心理学家纽厄尔（Newell）和西蒙（Simon）等人模拟了人用数理逻辑证明定理的思维过程和步骤，建立了机器证明数学定理的启发式搜索法，并编制了一个称为“逻辑理论家”的程序（简称 LT）。他们用这一程序证明了罗素和怀特的数学名著《数学原理》一书中第二章的 52 条定理中的 38 条（后来，美籍华裔数理逻辑学家王浩发明了一个王氏算法，证明了全部定理），这是用计算机对人的高级思维活动进行研究的第一项奠基性工作，也是人工智能的真正开端。同年美国国际商业公司（IBM）的塞缪尔（Samuel）利用对策论和启发式搜索技术编制出跳棋程序。这个“跳棋机”不是根据事先编好的一套固定的程序工作，而是能够自适应、自学习下跳棋，不断积累经验，并能根据对方的走步，从许多可能的步数中选出一个

较好的走法。这台机器不仅在1959年击败了它的设计者，而且还在1962年击败了美国一个州的跳棋冠军。这虽然是一个智力游戏，但却是早期一个较好的人工智能研究课题。同一时期美国麦卡锡（Macarthy）等人发起召开第一次以"人工智能的学习"为内容的讨论会，探讨人类各种学习与其他智能特征的基础，并研究如何在原理上进行精确的描述，以便用计算机进行模拟。这是人工智能发展史上的一个创始性聚会，从此正式使用"人工智能"这一术语，它标志着"人工智能"这门新学科的正式诞生。1963年，美国数学家鲁滨逊（Robinson）提出了归结原理，使机器证明获得了重大突破。最有说服力的莫过于1976年美国数学家阿佩尔（Apple）等三人共同合作借助于数理逻辑形式化的方法，完成了"四色猜想"的证明，解决了数学史上悬置了一百多年的著名难题。我国当代数学家吴文俊独树一帜，在机器证明领域内树立了意义深远的里程碑。他从构造性数学的观点出发，借助于坐标系将几何问题化为代数问题，发明了一种全新的方法，除在计算机上证明了非欧几何、圆几何、线几何及微分几何的全部定理外，还发现了过去不被人知的新定理，并进而给出了几何定理的机械化判定方法，被称为吴方法。它标志着机器证明和发现艰深的数学定理正变为现实可行，引起了国际数学界的瞩目。数学思维方法是人类思维的最高级的形式，机器证明为它赋予了机械化的可能。

机器人。机器人是机器智能的前沿。作为人工智能各种研究课题的综合运用，机器人的发展也极为迅速，从机器人手臂的最佳移动到实现机器人目标的动作序列的规划方法，无所不包，但最重要的一种是智能机器人。这种机器人的"大脑"是一部电子计算机，在存储器中放着程序和资料，使它能够在实践中不断增长才干和积累经验。它的"眼"是摄像机，能辨出所识别的目标。它的"耳"是一个话筒式录音机。它的"嘴"是一部扬声器。所以智能机器人具有感觉和理解周围环境，使用语言、推理、规划与操纵工具的技能，即能看得见，听得着，会思考，能说话，并且能通过学习适应环境，在不同条件下，采取相应的策略来完成自己的任务。

1986年，中国科学院在沈阳建立机器人示范工程，1990年8月建成。该工程以开发特殊环境下使用的第二代机器人产品为起点，不仅成为研究高级机器人的基础产品基地，还可以在高级机器人的人才培训、机器人性能测试、技术交流等方面发挥作用。由于机器人研制工作和工程建设同步进行，故示范工程做到边建设，边出成果，边见效益。他们瞄准世界机器人技术前沿课题研究开发，先后完成七台水下机器人的研制工作。第一台已在南海为海上石油钻探服务，第二台销往美国，第三台在成功地解决小丰满水坝拦污栅检查困难后，即将为日本打捞沉船服务。另外几台小型水下机器人已进入国内市场。我国第三代

机器人的研究工作也接近世界水平。此外，华东化工学院首次采用二维定性信息和三维定量信息相结合的识别方法，研制成功视觉智能控制的机器人手眼系统，使摄像机在很大范围内任意移动时，都能快速、准确地识别形状变化很大的各种工作，其定位准确性误差仅一毫米左右。视觉控制的智能机器人手眼系统是我国“七五”国防重点课题，也是自动化领域和信息处理领域的前沿课题。

专家系统。人工智能获得的又一成果是“专家系统”，这是一个以知识为基础的计算机系统。专家系统是把专家的知识分成事实和规则两部分，并以适当的形式存入计算机，建立起知识库，以知识库为基础，采用合适的“产生式系统”对输入的原始数据选择合适的规则，进行推理、演绎，作出判断和决策，起到有关专家的作用。因为专家系统是一个具有大量专门知识的计算机程序系统，可以在一定意义上成为那些优秀专家的化身，同时可在任何需要他们的地方出现，它能存储同一领域不同专家的知识，从而远远超出任何个别专家的水平。目前已研制的各种专家系统已广泛应用于地质勘探、医疗诊断、交通、教育、经济、法律等领域，并已收到较显著的效果。

总之，人工智能的重大意义在于，它把计算机应用提高到一个更高的新阶段，为计算机的应用拓广了领域。目前，美、日等国正在为研制高级的智能机而展开激烈的竞争，模糊数学的诞生为研制工作提供了理论基础。□

2 自然语言的机器翻译*

一　机器翻译的由来和含义

人的语言即自然语言包括书面语言和口语两种。人们靠这种自然语言进行读、写、听、说等活动。将一种自然语言翻译为另一种自然语言，是人工智能早期研究的活跃领域之一。

世界上的科技知识，20 世纪 70 年代每十年增长一倍，80 年代每五年增长一倍。科技知识这种急剧增长的趋势，有人形象地比喻为“知识爆炸”。显然传统的人工翻译方法面临一场严峻的挑战，电子计算机的问世和被广泛利用，为解决这个问题带来了新的曙光。

远在 1946 年，当电子计算机问世时，人们便想到利用计算机把一种语言文字翻译成另一种语言文字，这就是所谓“机器翻译”。我国科学家对机器翻译进行了大胆探索，在英汉、日汉、俄汉翻译系统方面有了长足的进展。

最初，人们认为翻译只包括查词典和语法分析，但事与愿违，使用这种方法并没有达到预期的效果。原因是当一个人听语言时，不仅是依赖于他的语法知识，而且还要运用与他所讨论的内容有关的知识。因为某个词和句子在上下文中有一种特定联系的意义（事实上，上下文理解是自然语言理解的基础，也是机器翻译的基础）。早期的

* 本文原载于《自动化博览》1993 年第 6 期。

机器翻译系统未获得成功是因为没有去尝试理解它所翻译的内容究竟是什么，所以机器输出的新语言不能精确地复原原语言的意义。从60年代中期起，同自然语言处理有关的研究大都转向回答系统。在规定的范围内，机器理解和回答人们所提出的问题，依据机器对人所提问题的回答，可以判断机器是否理解了人所说的语言。1972年，美国学者温诺格拉特在麻省理工学院提出了一个著名的程序，叫SHRDLU系统，这个系统仿照人主要靠语言、语义和概念理解自然语言的规律，在计算机上模拟机器人搬运木块的回答程序。温诺格拉特认为，人是运用自己的全部知识和智能来理解语言的，他提出的这个程序就是以语法、语义和推理以及上下文和内容知识相结合的一个具体程序，用以成功地进行了人同计算机灵活地对话。这是自然语言处理领域的一个重大突破，震动了当时的人工智能界。自此以后，关于"自然语言的理解"的研究受到各主要国家的重视，并且得到了发展。于是，研究翻译自动化的一门新学科——机器翻译诞生了。其任务是研究从一种自然语言到另一种自然语言的翻译规则，建立起适合于电子计算机进行工作的翻译规则系统。

翻译是一个相当复杂的人脑思维活动，机器翻译是由计算机按一定程序模拟人脑思维的高级智能活动，所以属人工智能的范畴。机器翻译不仅要有一整套语言的分析规则系统（或称理解系统），而且还包括另一种语言的生成系统（或称表达系统），所以，机器翻译是一种高级的人工智能，它是语言学、数学、计算机技术、自动化技术等多种学科相互结合、渗透的产物，是一门新兴的边缘性交叉学科。

二　机器翻译的基本原理

简单地说，利用机器进行翻译，就是让机器模拟人的翻译过程。人在翻译之前必须掌握两种语言的词汇和语法，机器翻译也是这样，它在翻译之前就已经在它的存储器中存入语言工作者编好的并由数学工作者加工过的机器词典和机器语法。人翻译所经过的步骤或过程，机器也同样进行。如先查词典，得到词的意义和一些基本的语法特征（如词类等），如果查词典得到的不止一种意义而是多义字（或词），那么，就要根据上下文来选取所需要的一种。在词汇和基本语法特征弄清楚之后，就要进一步明确各个词之间的关系，此后再根据语法的要求组成包括改变词序、翻译原文词的一些形态特征及修辞在内的译文。

三　机器翻译的过程

机器翻译系统是一个庞大的应用性软件系统，一般由原文输入、原文分析、译文综合和译文输出四个部分组成，这就是机器翻译的过程或步骤。下面以英汉翻译为例，作简要说明。

1. 原文输入：原文字母和符号必须按一定的编码规则转换成计算机可以接受的二进制数字输入计算机。

2. 原文分析：包括查词典和语法分析两个阶段。关于前者，通过查词典，给出词或词组的译文代码和语法信息，为以后的语法分析和译文的输出准备条件。通过查词典，原文中的词在语法类别上便可成为单功能的词，在语义上成为单义词（某些介词和连词除外）。接着就进入语法分析阶段。语法分析的任务是：① 进一步明确某些词的形态特征；② 切分句子；③ 找出词与词之间句法上的联系，同时得出英汉语的中介成分，为下一步译文综合做好准备。

3. 译文综合：原文分析是整个翻译规则系统中最复杂的部分，译文分析，相对来说则比较简单。这一阶段的第一个任务是把应该移位的成分调动一下。第二个任务是修辞加工，即根据修辞的要求增或删一些词。第三个任务是查汉文词典，根据译文代码找出汉字的代码。

4. 译文输出：通过汉字输出装置将汉字代码转换成文字，打印出译文来。

四　目前存在的问题

自然语言的人译与机译根本不同之处是，人在完成输入、查词典和原文分析后，经过一系列的思维过程，然后从语法和意义上进行转换翻译，形成原文输出。人具有联想、理解、判断能力，而机器只具有形式逻辑比较能力，或者说，计算机只能按人们事先规定好的一套规则进行工作。因此，要使机器翻译能够达到人译的水平，就必须具备联想处理的能力，这是目前的冯·诺伊曼型计算机所不能胜任的。现在世界上许多国家正在研制新一代计算机，它在系统结构方面的目标是超越冯·诺伊曼型计算机。它的主要功能将着重于知识信息处理，具有逻辑推理能力，可用语言、文字、图形等多种多样的形式直接进行人机通信，将广泛应用于自然语言理解、机器翻译、机器证明、各种模式识别等。

自然语言不仅是人工智能所要处理的核心问题，同时也是机器翻译所面临

的难题。只有研制出能够理解和表达自然语言的机器智能系统,才能进行理想的机器翻译。从这个意义上说,机器翻译的成功实现,将是人工智能的一个重大突破。我们相信,语言文字工作者、人工智能和机器翻译研究工作者,协同努力,共同攻关,必将为机器翻译开拓新的前景。□

3 模式识别及其应用*

所谓模式(pattern)，指的是人脑中存储的有关讨论对象的标准样本。人在同外界接触时，有意识地从头脑里的样本库中提取样本，与待识别对象的特征相比较，按标准样本将对象归类，这种思维过程就是模式识别。每个人的头脑里都存储有自己曾接触过的人和物的形象，这些就是样本。有人叩门来访，你觉得来人似曾相识又一时记不起来，于是开动脑筋，将来人的容貌与头脑中的样本相比较，终于认出了故人。这就是一种模式识别。实际上，人们在生活中无时无刻不在用感官通过大脑对事物进行模式识别。为了识别事物，人们总要依据一定的标准对它们进行分类。在现实世界中，有许多事物可以依据精确标准把它们分为界限分明的类别，每个事物要么属于某一类，要么不属于该类，非此即彼，明确肯定。我们把事物有明确类属的这种特性称为清晰性，把这类事物称为清晰事物。但是，对于另外一些事物，我们却无法找到精确的分类标准，关于某一事物是否属于某一类，很难作出明确肯定的断言。老年人、中年人等就是这种事物类。由于这种事物从属于某一类到不属于该类是逐步过渡而非突然改变的，又由于同类别之间不存在截然分明的界限，因而不同人对同一事物可能作出不同的归类。我们把事物这种类属的不清晰性称为模糊性，把这类事物称为模糊事物。

标准样本是按照事物的一定性态对事物进行划分而建立起来的。按照清晰性态确定的是清晰样本，按照模糊

*本文完成于 1995 年。

性态确定的是模糊样本。被识别对象总是非标准的，否则就用不着识别。按清晰样本进行识别是精确思维的内容，对象是标准样本的某种变形，经过适当处理，可化为标准形式进行识别，作出非此即彼的结论，这种模式识别叫做清晰识别。但大多数模式识别是按模糊样本进行的，如预报天气，诊断病情，辨认口音、字迹、容貌及其他图形等，多为模糊识别，不但样本具有模糊性，识别的规则和方法也是模糊的。前一类模式识别可以编成严格的程序，化为一系列精确的计算，由机器来执行。但模糊识别的机制要复杂得多，尚待进一步探索。

从广义上说，模式识别是用机器来分析、识别人类所具有的感觉（视觉、听觉、触觉、味觉、嗅觉）和思维（理性的和逻辑的），而在当前主要是人类视觉和听觉能力的模拟。视觉和听觉是人接受外界信息的主要通道，通过视觉通道取得的信息是色彩、形状、空间等各种形式的图像，通过听觉通道取得的信息对于各种音响、视觉图像识别有特别重大的意义。从信息论的角度来看，图像所包含的信息量最大，不仅有灰度还有色彩，不仅有平面还有立体等，所以其内容极为广泛。实验表明，人有 80%以上的信息是来自视觉，有 10%以上的信息是来自听觉。所以具体地说，模式识别是利用机器对物体、图像、语言、字符等信息模式进行自动识别。电子计算机具有高速运算、大容量存储和逻辑判断、数据处理等功能，所以电子计算机是代替人作模式识别的一种理想机器，它既节约人力，又细致、耐心、迅速和准确。遵循模式识别的原理，设计出软件系统，经过硬件实现。模拟人的视觉能力就是用计算机来做图像的识别和理解工作，模拟人的听觉能力就是用计算机来做语言（或各种声音）的识别和理解工作。识别和理解代表不同的涵义，理解比识别更深了一步。当前，模式识别技术所能做到的基本上是一种分类工作，离“理解”还有相当大的距离。尽管如此，模式识别技术并不是对人的能力的简单代替，在很多场合下，它可以完成个别人的能力所无法做到的工作，它已成为计算机科学中的一个重要应用领域，是在自动控制理论、计算机科学、自动化应用等基础上发展起来的多学科交叉渗透的边缘性学科，是人工智能研究中迫切需要解决的课题。

20 世纪 60 年代是模式识别技术的发展初期，人们不断探索和开拓它的应用领域，并在高能物理的图片识别方面取得了成果。70 年代初，随着遥感技术的发展和地球资源卫星的发射，利用图像处理、模式识别技术来分析所获取的信息，以了解地球资源分布情况方面取得了令人满意的结果。近年来，模式识别在地球资源和人类环境、生物医药工程、生产过程自动化、智能机器人、文件处理与管理自动化、军事、指纹与足迹的识别以及法庭审判与商业自动化等方面都得到了卓有成效的应用，但这些均属视觉方面的应用。在听觉、语音方面也有实际的应用。近年来，由于微电子学的发展，人们已经使计算机具有初步

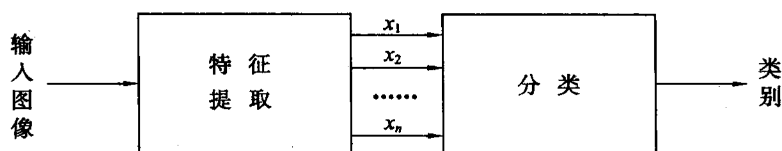
的听话和说话的能力，并且有望将来使用计算机就像用普通的电话那样简单和方便。1980年初美国伊利诺伊州已采用能听懂人的语言的计算机为打电话者接通长途。这种听懂人的声音并能作出回答的计算机的出现，意味着人们又向未来的目标迈出了坚实的一步。

人对事物的认识还是很初步的，依据这些初步的认识，我们可以看到人对无限发展变化着的认识是有指向性的，并且是有选择地抽出事物属性中的主要特征加以认识。通过感官人们从客观事物获得大量的杂乱的输入信息，经过思维活动将这些信息（数据）转换、归纳、划分为若干类别的输出信息，判断客观事物应属于哪一类。识别过程就是把知觉的模式与预先通过学习或经验记忆的模式进行核对的模式匹配过程。模式是具有多维性质的信息结构或信号，一般无法用数学方法来描述，是属于非数值问题。

模式识别方法可分为两大类：清晰模式识别方法和模糊模式识别方法。

1. 清晰模式识别方法

（1）统计法。建立在概率论基础上的模式识别方法，叫做统计模式识别方法，或简称统计法。这种方法最终归结为分类问题，它是模式识别中应用较广的一种方法，其基本做法是首先从待识别模式中抽取特征参数，即着眼于找出能反映图像特点的特征度量，把图像的数据进行信息压缩，以抽取特征参数。如果抽取 n 个特征，能够基本上描绘原来的图像，则图像就可以用 n 维向量来表示。因此对于图像分类，就相当于把特征空间划分为若干个部分。当输入一个图像时，就根据相应的特征向量属于特征空间中的哪一部分，而决定属于哪一类。其原理如下图所示。



简言之，统计法就是抽取特征，与机器中存储的模式进行匹配，匹配得上，就能判定类别。这类似于一个人见到阔别多年的老朋友时，总是通过把看到的老朋友当前的特征与过去记忆中的特征进行比较来认识。

（2）语言结构法，或简称语言法，也称几何法。人的认识并不是赤裸裸的，而是要借助于语言。人的认识还是一个思维过程。语言是认识的物质手段，是思维的物质外壳。识别图形可以认为是用语言描述图形的过程，能识别的对象都可以用语言描述清楚。语言结构法就是基于人在认识过程中视觉和语言的联系而建立的，即通过用语言描述对象结构的相互关系来识别对象。众所周

知，任何一种语言都是由一些词、短语等按一定的语法组成的。句子由短语组成，短语由单词组成，最基本的元素是单词。语法规定着单词、短语、句子之间的结构关系。人们由此而得到启发，把图像分解成一些直线、斜线、折线、点、弧等基本元素，剖析这些基本元素，研究它们是按怎样的规则（语法）构成图像的，即从结构入手，检查待识别图像属于哪一类"句型"，是否符合事先规定的"语法"。按照这个原则，若语法正确，图像就能被识别出来。

采用语言法可以清晰地描述一些结构复杂的模式，有了这种对目标的描述，就可用机器中已有对分类规则的描述进行识别。乍看起来这种方法比较麻烦，但实际上使用计算机来处理就很容易实现。

2. 模糊模式识别方法

人类生存的环境基本上是一个模糊环境，人们在生存活动中，经常接触各种模糊事物，接受各种模糊信息，随时要对模糊事物进行识别。客观事物的特征带有不同程度的模糊性，大多数模式识别是按模糊样本进行的。人脑能作清晰的模式识别，更善于作模糊的模式识别。人脑思维的低精度特点，在模式识别中成了得天独厚的优点。由于精确度低，识别过程不必以获取大量精确数据为前提，只需依据少量模糊信息，即可作出足够准确的识别，因而适用范围广。人脑是一个极其复杂的网络系统，人脑识别是能动的、灵活的，善于区分对象的主要特征和次要特征、稳定的特征和易变的特征，估计对象可能的变化或有意制造的假象。人脑的这些功能都与模糊性有关。由于人脑思维的不确定性特征主要是模糊性的而不是概率性的，统计模式识别与人脑的模式识别在机理上有重大差别，正在发展的以模糊数学为基础的模糊模式识别理论，将为理解人脑模式识别的机理提供有价值的材料。我们就模糊数学中有关模糊模式识别加以论述，以考察人脑识别活动的特点。模糊模式识别按对象可分为个体识别与群体识别两类。

（1）个体识别。当问题是论域 U 上所有作为标准样本的模糊集合已经确定，要求识别论域 U 中某个具体对象归属哪个样本类时，是个体识别。如手写字的识别，手写的汉字与标准印刷体的汉字是有区别的。同一汉字，由不同的人手写，其大小、形体是不同的，千姿百态，构成汉字的基本笔画具有模糊性。一个汉字代表一个模糊偏序结构，每一个标准印刷体汉字都是一个样本。给定一个手写汉字，对于每一个样本都有一定的隶属度，其中必有最大者，据之决定该手写汉字隶属的样本（英文手写字母的识别也有类似的问题，26 个英文字母是 26 个样本，各代表一个模糊集合），这即为最大隶属原则。

最大隶属原则　设$\underset{\sim}{A}_1, \underset{\sim}{A}_2, \cdots, \underset{\sim}{A}_n$是论域 U 上的 n 个模糊子集，u_0 是 U 上

的固定元素，u_0 为待识别对象，u_0 对$\underset{\sim}{A}_i$的隶属度记为$\underset{\sim}{A}_i(u_0)(i=1,2,\cdots,n)$。如果存在 $j(j=1,2,\cdots,n)$，使得

$$\underset{\sim}{A}_j(u_0)=\max\{\underset{\sim}{A}_1(u_0),\cdots,\underset{\sim}{A}_n(u_0)\},$$

则认为 u_0 相对隶属于$\underset{\sim}{A}_j$，即元素 u_0 应归属模式$\underset{\sim}{A}_j$。

这种直接由计算元素的隶属函数值来判别具体对象归属谁的方法，又称为模糊模式识别的直接方法。这种方法适合处理具有如下特点的问题：① 用来比较的模式是模糊的；② 待识别的对象是确定的。

直接识别方法是模拟人脑处理问题的思维过程而构造的一种数学模式，它的形式和归类方法虽然很简单，但不能有效地处理许多问题，如机器自动识别染色体或白血球分类等课题中，常把问题归结为一些简单的几何图形的识别。但具体图形在现实生活中都带有不同程度的模糊性，不可能绝对标准，总会带有不同程度的误差。

（2）群体识别。直接识别法所要识别的对象是单个情况，但是在现实生活中，有时要识别的对象并不是论域上的单个确定的元素，而是论域上的模糊子集，即考虑两个模糊集合靠近程度（或叫贴近度），这时直接识别方法便失去效用。对这类模糊识别是将待识别的模糊集合与所有作为样本的模糊集合进行比较，确定它与哪个样本最贴近，就认为它相对隶属于最贴近的那个样本。这叫做择近原则。例如工业生产中，未经处理的毛坯零件是模糊的，对其摄像获得的图像更加模糊，要从图像中识别所需要的零件，就得用择近原则。以下说明此原则。

定义 1 若$\underset{\sim}{A}$与$\underset{\sim}{B}$为论域 U 上的两个模糊子集，则$\underset{\sim}{A}\circ\underset{\sim}{B}=\bigvee_{u\in U}[(\underset{\sim}{A}(u)\wedge\underset{\sim}{B}(u)]$称为$\underset{\sim}{A}$和$\underset{\sim}{B}$在 U 上的内积，$\underset{\sim}{A}\odot\underset{\sim}{B}=\bigwedge_{u\in U}[A(u)\vee\underset{\sim}{B}(u)]$称为$\underset{\sim}{A}$和$\underset{\sim}{B}$在 U 上的外积。符号"$\wedge$"为取极小运算，"$\vee$"为取极大运算。

定义 2 若$\underset{\sim}{A}$和$\underset{\sim}{B}$都是 U 上的模糊子集，$\underset{\sim}{A}$与$\underset{\sim}{B}$的贴近度记为$(\underset{\sim}{A},\underset{\sim}{B})$，则定义$(\underset{\sim}{A},\underset{\sim}{B})=\frac{1}{2}[\underset{\sim}{A}\circ\underset{\sim}{B}+(1-\underset{\sim}{A}\odot\underset{\sim}{B})]$。

择近原则 若$\underset{\sim}{A}_1,\underset{\sim}{A}_2,\cdots,\underset{\sim}{A}_n$ 为论域 U 上的模糊子集，$\underset{\sim}{B}$为 U 上待识别的模糊子集，若有 $j(j=1,2,\cdots,n)$，使得

$$(\underset{\sim}{B},\underset{\sim}{A}_j)=\max_{1\leqslant i\leqslant n}\{(\underset{\sim}{B},\underset{\sim}{A}_i)\},$$

则认为$\underset{\sim}{B}$与$\underset{\sim}{A}_j$最贴近，而应把$\underset{\sim}{B}$归于模式$\underset{\sim}{A}_j$。这个方法又称为模糊模式识别的间接方法。

对模式识别的展望。因为模式识别技术是用计算机模拟人的各种识别能力，所以当今的模式识别技术都是推论性质的。用推论的方法进行识别，速度

慢，同时也不能保证识别的准确性。为了改进推论性质的识别技术，在模式识别中，不仅要“识别”，而且还要“理解”，在“理解”的基础上进行“识别”。从人工智能的观点来看，“识别”是对客观事物按其物理特征来进行分类，“理解”是指通过知识工程中的推理技术对客观事物内涵的了解。要使机器具有视觉和听觉，首先必须要求它能对复杂输入数据进行有效处理使其“理解”这些输入，而要理解则要求具有大量有关感受到的事物的基础知识。目前，模式识别技术所能做到的基本上还只是一种分类工作，离“理解”还有较大的距离。因此，从总体上说，模式识别现尚处于初级阶段。模式识别既涉及抽象思维，也涉及形象思维，图形识别就是形象思维的职能。钱学森指出：“我们不清楚形象思维的规律，就是图形的识别也还是个大问题，不知道人脑是怎样识别图形的。”目前，如何结合人脑的研究，使计算机带有接近人类的智能，从而使模式识别系统更加实用可靠，以感知为基础的模式识别学科如何与以知识为基础的人工智能学科相结合，无疑是今后的战略研究方向。□

4 定理机器证明*

一 定理机器证明的缘起和意义

数学，不论是学习还是创新，最耗时费力的劳动往往消耗在定理的证明上，而不是在真理的发明、发现上。数学是一门严谨的科学，证明是完全必要的，证明的严密性也是完全必需的。因此，数学工作者爱好证明，重视证明，但也常常为证明之艰难而困扰着。在数学史上有许多欲证而不能的猜想，如 18 世纪德国数学家哥德巴赫(Goldbach)提出的猜想，即所谓哥德巴赫猜想。通过严格的数学来论证，并被证明了的猜想，就直接变成定理。下面谈的"四色猜想"就是著名的一例。

17 世纪，德国数学家、哲学家莱布尼茨提出一个宏伟设想：制造一种推理机器，把人类关于理性的归纳和演绎机械化。但在莱布尼茨时代，由于生产力发展水平不高，这种以机器模拟人脑思维活动的研究没能得到迅速发展，因而推理机没有被制造出来。但莱布尼茨思想的价值仍存在，这就是对以后的布尔代数和数理逻辑起到了促进和完善作用，也是以后电子计算机创立的一种"催化剂"。1937 年，英国数学家图灵(Turing)也发表了《理想自动机》的论文，该文给出了可计算性这一概念的严格的数学

* 本文原载于《知识工程》1992 年第 1 期，后由《自然辩证法研究》1994 年第 5 期、中国人民大学《自然辩证法》1994 年第 7 期转载。

定义，并论证了任何需要精确地加以确定的计算过程，都能由图灵机完成。为人们清晰地描绘出理想自动机的蓝图，为以后电子计算机的创立奠定了基础。经过以后几代人的艰苦努力，美籍匈牙利数学家冯·诺伊曼（Von Neumann）等人在图灵提出的一般机器的概念的启迪下，终于在1946年研制出了世界上第一台电子数字积分计算机，这是人类历史上具有划时代意义的大事。

计算机，顾名思义，是一种计算的机器，具有快速运算的能力。它首先使人们从繁琐复杂而又十分单调的四则运算劳动中解放出来，从而结束了几千年来数学家只是用一支笔和一张纸进行研究的手工式劳动，能把大量的时间用在真理的发现和发明上；其次，由于电子计算机又具有逻辑推理和判断功能以及存储功能，对某些艰深的定理的证明，可以借助于计算机来实现，这对于科学，特别是对于数学发展的影响是难以估量的。如同望远镜之于天文学、显微镜之于生物学打开了人类认识自然的奥妙一样，计算机和计算机科学的发展为数学的发展开拓了新的天地。它不仅使一些长期因计算量浩大而不可设想的问题有了解决的可能，且某些或某类数学定理的证明，也可以在计算机上来实现，即把质的困难转化成量的复杂，而后者对于计算机来说是轻而易举的，因而使得定理的证明化难为易，这就是定理机器证明，也称为定理的机械化或自动化证明。

证明定理是人类特有的智能行为，是逻辑演算的过程。定理机器证明就是人类证明定理的过程通过一套符号体系加以形式化，变成一系列能在计算机上自动实现的符号演算过程，也即把具有智能特点的推理演算过程机械化。

用机器证明数学定理的想法最初是由莱布尼茨提出来的。20世纪70年代，德国数学家希尔伯特（Hilbert）创立了数理逻辑，为机器证明定理提供了理论和方法。另一方面，电子计算机的问世和发展，又为机器证明定理提供了物质条件。因此，证明机械化的研究蓬勃发展起来。1950年，波兰数学家塔斯基（Tarski）提出并证明，一切初等几何和初等代数范围内的命题都可以用机械化方法判定。令人惋惜的是，塔斯基的方法和后来其他人改进的都极为繁复而难以实现，但它对后来的影响是巨大的。以后数理逻辑的发展和完善为机器证明定理提供了一种强有力的描述语言。1956年，美国的西蒙（Simon）、纽维尔（Newell）和肖乌（Shaw）三人通过研究证明定理的心理过程，提出了一个称为“逻辑理论家”的程序（LT）。这个程序不是刻板的固定算法程序，而是模拟了人用数理逻辑证明定理时的思维过程、规则和所采取的策略、技巧以及简化的步骤。它在计算机中先存储一些公理，然后用分解、代入和替换等方法来处理待证明的问题，如果这些子问题最终能变换成已知的公理或已证明过的定理形式，那么这个问题就得证了。西蒙等人利用这一程序，在大型计算机上证明了罗素和怀特海的名著《数学原理》第二章中的全部定理。1956年，美国数学家鲁

滨逊(Robinson)提出归结原理，这一原理与西蒙等人的证法相反，他的基本出发点是要证明任何一个命题为真，都可以通过证明其否定为假来得到。归结原理的提出使机器证明获得了某些突破。虽然不少人在计算机上加以实现，但由于时间和空间的限制，这种理论精美的方法，终不能达到实用的程度。值得提出的是，20 世纪 50 年代末美籍中国学者、数理逻辑学家王浩发明了“王浩算法”。他把整个机器证明过程规则化，整个系统由十七条规则及三个主要过程 P、Q_AE、Q 组成。过程 P 用于处理命题运算，过程 Q_AE 用于处理最小辖区域形中的 AE 谓词演算，Q 用于处理一阶逻辑运算。王浩于 1958 年在 IBM704 机上，用 3 分钟时间证明了《数学原理》一书前五章中的约 220 个命题演算定理。1959 年，他在 9 分钟的时间内完成了《数学原理》中全部定理(350 条以上)的证明。我国数学家吴文俊认为，王浩在机器证明上获得了突破性成果，其有关机器证明的一些精辟论点，诸如以量的复杂取代质的困难，以及基础机证与特例机证应有区别等论点，更是发人深思。但这些方法的弊端是，它难以保证证明全部定理。

二　四色问题和吴文俊方法的出现

古代的地图大都是五颜六色的，使人看了眼花缭乱。能否减少颜色直至四种就可把不同国家或地区的边界区别开来？这就是著名的“四色问题”，也称为“四色猜想”。它最早是由德国数学家麦比乌斯(Möbius)于 1840 年提出来的，是图论也是拓扑学中一百多年来一直困扰人们的著名难题。尽管它的严格的数学证明很艰难，但陈述却简洁易懂：对于平面或球面(如地图仪)上任何地图的涂色中，不管有多少个国家或地区，也不管它们怎样设置，至多只用四种颜色对地图涂色，就可以使相邻的国家或地区着有不同的颜色。只用三种颜色有时显然是不够的，如欧洲的卢森堡，被德、法、比三国所包围，没有四种颜色是分不清的。但可以证明，五种颜色一定够用，这就是五色定理。问题是四种颜色是否够用，这就是有名的四色问题。

用四种颜色对地图着色，虽是一种猜想，但到目前为止，尚未发现有四种以上颜色涂色的现代地图，可是它的严格的数学证明却是繁琐而艰难的。许多数学家(如德国的闵可夫斯基)对它的难度估计得并不充分，虽认真研究了四色问题，试图用数学方法来加以证明，结果无一不是以失败而告终。直到 1874 年，美国数学肯普(Kempe)才给出一个精致巧妙的证明方法，证明“四色”的答案是正确的，但遗憾的是 1890 年，数学家海伍德(Heawood)发现肯普的推理有一个

漏洞，其结论并不成立。后来有许多人设法弥补那个漏洞，但没有一个人能成功。虽然如此，肯普方法的价值仍旧存在，这就是他的思考路线带给了我们一些成果。成果之一就是海伍德利用肯普的思考路线证明了对所有的地图，五种颜色是足够的，这就是上面所说的五色定理。但仍没有判断出四色猜想的真伪。到了 20 世纪 70 年代，某些有远见的数学家指出，四色猜想是一个正确的定理，但它需要一种数学中前所未有的方法来证明，这个方法是不能单靠人力所能胜任的。石破天惊，证明奇迹般地出现了。1976 年，美国伊利诺斯大学的青年数学家阿佩尔（Apple）、哈肯（Haken）和计算机专家莫尔（More）共同合作，他们依从肯普的思考路线，根据一些不可避免的图形，将四色问题分为 2 300 多类（这种分类比肯普的分类细致得多，而且每一分类都很复杂），借助于数理逻辑使证明形式化的方法，利用计算机按照他们所设计的程序，用了 1 200 个小时的机器时间，做了近 100 亿个逻辑判断，从而完成了四色猜想的证明，终于使四色猜想变成了四色定理。悬置了 120 多年未能解决的著名数学难题得到了肯定的回答。如此繁重的计算工作量和如此之多的逻辑判断，如果不是依靠计算机所具有的高速度运算和大容量存储的功能，光靠数学家几千年来传统的手工式劳动来完成，据估算，大约得需要 30 万年。这简直是一个不可思议的大难题。因此，四色问题的解决，被列为 1976 年科学上的重大研究成果。但这一成果重大意义还不仅在于四色问题本身的解决，还因为这个方法的本身对数学中的很多难题，突破了数学理论传统的证明方法，开创了人和机器合作去解决问题的新途径，从而导致了“定理机器证明”这一崭新思想方法的确立。

定理机器证明本身的艰难是显而易见的。《数学原理》中 300 多条定理用计算机证明显然不在话下，但即使是四色猜想，也只能是计算机辅助证明。这种“一理一证”的极端方式不应视为定理机器证明的主流。定理机器证明应极大限度地兼顾通用性。因此，定理机器证明的进一步发展必须要开创全新的局面。

机器证明思想在我国古代已有萌芽，中国的古代数学基本上是一种机械化的数学。例如，汉初完成的名著《九章算术》中，对开平方、开立方的机械化过程就有许多说明。在开平方法中是借用一根算筹来表示未知量的平方，在开立方法中借用一根算筹来表示未知量的立方，这就给所列出的筹式一个代数方程的意义，开平方或开立方的各个演算步骤也就是解方程、求正根的过程。在宋元时代，我国更创立了“天元术”，引进了相当于现代未知数概念的天元、地元、人元、物元，把许多问题，特别是几何问题转化成代数方程与方程组的求解问题。这个方法用在几何上就相当于几何的代数化。12 世纪，杨辉、李冶、朱世杰又引进了相当于现代多项式的概念，建立了多项式的运算法则和消元法有关的代数

工具，使几何代数化的方法得到了系统的发展。几何的代数化实际上就是现代解析几何的前身。我们知道，法国数学家笛卡儿17世纪创立的解析几何，确实能使几何问题代数化，即用坐标法把几何问题化为代数问题，但并没有实现几何定理的机械化证明。我国数学家吴文俊在深入研究中国古代代数，特别是在研究《九章算术》的基础上，提出了自己独具特色的机械化判定方法，并于1977年在《中国科学》第6期上发表题为《初等几何判定问题与机械化证明》的文章，又于1984年出版了世界上第一部关于几何定理机器证明的学术专著《几何定理机器证明的基本原理（初等几何部分）》，该书成为机器证明领域内的里程碑。吴文俊用他自己建立的机器证明方法和编写的程序，在计算机上成功地证明了非欧几何、圆几何、线几何以及微分几何中的600多条重要定理。这些定理大多是不平凡的，其中还有些是过去不被人知的新定理，而证明每条定理所用的机器时间才几秒钟。吴文俊的方法不胫而走，受到了国际学术界的高度重视并引起了强烈的反响，被誉为“吴方法”。它的出现标志着机器证明定理和发现艰深而不平凡的新定理正变为现实可行，从而结束了具有数千年历史的传统的初等几何定理的证明方式。今后，如果人们在初等几何范围内提出新命题而不知其真伪时，只要把该命题化成代数问题，用吴方法编写程序，上机一试便知分晓，从而使数学家们得以摆脱繁琐的手工式劳动（计算和证明），腾出时间和精力，把聪明才智用到真正创造性的工作中去，这必将大大推动数学的发展。

为了研究和学习“吴方法”，推动“吴方法”的发展，在国家有关科技部门的支持下，中国科学院系统科学研究所于1990年成立了以吴文俊为核心的中国数学机械化研究中心。该中心不仅在理论和方法上充实了“吴方法”，而且在多项式领域、非线性规划、机器人学、控制论和几何模型等应用研究方面也获得了具有国际水平的成果。特别是在自动推理领域，“吴方法”具有划时代的意义，是近十年中自动推理领域最激动人心的突破性进展。

欧美一些国家竞相研究“吴方法”，并邀请吴文俊前往讲学。如20世纪80年代初期，美国数学会邀请吴文俊出席在科罗拉多举行的机器证明会议，会上对“吴方法”给予了极高的评价。会议记录《机器证明25年》一书，有四分之一的篇幅是关于吴文俊方法的论文。1984年，美国《自动推理杂志》破例重印了吴文俊方法的奠基性论文《初等几何判定问题与机械化证明》。

三　吴文俊方法的基本思想简介

在数学里要求证明的问题，往往是由条件和结论两部分所组成。所谓数学

证明,众所周知,就是根据所给的条件,已有的定义、公理以及已被证明成立的定理,按照一定的推理格式导出结论。这在相当大的程度上要依赖于人们的智慧、才能和经验,并且需要较长的时间,致使人们望而却步。几何定理的证明更是这样。吴文俊认为,我国几何学的许多重要领域,例如微分几何与代数几何,往往一开始就假定了一个数系(一般是一个数域,甚至是一个特殊的实数域或复数域),由此构成仿射空间或投影空间,然后用坐标或以函数与导数间的关系来引进各种几何图形,如曲线、曲面及其几何关系等。看来,在数的基础上建立起整个几何学的体系已是大势所趋,但也不能不考虑几何学的直观背景与来源,以及它的基础问题。如何从原始的现实形象提炼出一套几何学的公理系统,又如何从公理系统发展成坐标系统,使几何定理的证明完全化成代数问题,使代数方法得以在几何上发挥作用,看起来是平凡的,但做起来并非轻而易举。首先,因为解决这些代数问题计算量过大,使人望而却步;其次,由于代表几何关系而出现的那些代数式往往杂乱无章,使人们不知所措。幸而现在出现了电子计算机,对繁杂的计算已经有了有效的处理办法。因此,如何把杂乱无章的代数关系处理得井然有序,使计算机得以在证明几何定理上发挥威力,便成为问题的关键所在。对此,吴文俊证明了如果一种几何的附属数系是一个数域,即乘法是可交换的,则假设与终结部分都可由多项式等式关系来表达的那类定理的证明都可以机械化。这一类定理称为等式型定理。它的机械化问题可归结为一个纯代数式的机械化问题,包括几何学中最主要的大部分定理。事实上,有序或无序的常用几何学、投影几何学、非欧几何学等,它们的几何定理都有机械化证法。因此,在理论上,这些几何学的定理证明,都可以借助于计算机来实现。对此,吴文俊已出版的《几何定理机器证明的基本原理(初等几何部分)》,把几何定理的机器证明分成以下三个步骤:

(1) 几何的代数化与坐标化。从几何的公理系统出发,引进数系统与坐标系统,使任意几何定理的证明问题成为纯代数问题。

(2) 几何的机械化。将几何定理假设部分的代数关系式进行整理,然后依确定步骤验证定理终结部分的代数关系式是否可以从假设部分已整理的代数关系式中推出。

(3) 用计算机作最后验证。依据第二步中所确定的步骤编成程序,并在计算机上实施,以得出定理是否成立的最后结论。

由于现在的计算机尚只能识别有限的事物,因此几何定理机器证明的一个先决条件是第二步中所确定的代数关系式必须是以有限的形式出现。例如这些代数关系式都是以多项式的形式出现,且其系数都是整数,则对于计算机的使用只是依据第二步中所确定的步骤编制程序的问题,不会有任何实质性的困

难。如果一门几何可以找到这样三个步骤（事实上只要前两个步骤即可）来完成定理的证明，我们就说这门几何可以机械化，并把可以机械化的这一结论称为机械化定理。由以上所述可以看出，由公理化到机械化大体经历了如下一个过程：

公理化→代数化→坐标化→机械化。

一门几何学能否机械化并不显然。按“吴方法”，几何的公理、定理与证明，实质上都可以用有限次的构造步骤来叙述以至完成，这也就为我们提供了几何定理证明有可能机械化的依据。吴文俊证明了：奠基于各种公理系统的每种初等几何学，只要相当于乘法交换律的某一公理成立（即乘法是可以交换的），则假设与终结部分都可以用多项式等式表达的那类定理的证明，大都可以机械化。这些几何定理的证明，可以借助于计算机来实现，这一类定理即前面说过的等式型定理。可以机械化的几何包括了多种有序或无序的常用几何学（即欧氏几何的那种初等几何）、投影几何、非欧几何与圆几何等。

以上说的几何定理证明的机械化，其寻求的是一般方法，不仅适用于个别的定理，而且适用于某一类型的定理，甚至可以说是某一种几何的所有定理。几何的代数化是定理机器证明的关键一步。

值得提出的是，“吴方法”不仅能证明几何定理，而且在数学其他领域或物理学科上也大有用场。美国数学家乌拉（Ulam）为了要探讨应用中广泛出现而现代数学又对之还无能为力的非线性现象，便是在计算机上进行试验，发现了一些规律。近年来，微分方程弧立子解的获得，就是首先在计算机上发现的，这是应用数学上的一个重大突破。吴文俊本人还曾利用他的方法，在开普勒定律之下成功地证明了著名的万有引力定律。最后我们引用著名的工程控制论专家、中国科学院院士宋健的话来结束本文：“吴文俊机器证明理论，为高度公理化的数学金字塔通向现代社会建立了一座桥梁，必将对社会经济、生活、生产以及科研产生不可估量的影响。”□

5 万能机器人*

一　机器人的含义、产生和发展

“Robot”一词最早出现于1920年捷克斯洛伐克剧作家查别克(K. Capek)的幻想剧《罗萨姆的万能机器人公司》中，以后科学家们采用了这个名词，并沿用至今。

实际上，在该剧中，“Robot”这个词的本意是苦力，指的是“人造的像人一样的机器奴隶”，是一种人造的劳动力(即人造人)。但是，现在机器人的含义与此迥然不同。例如1981年，东京第11届国际工业机器人讨论会上确认的定义为：工业Robot是用于搬运材料、零件、工具或专用设备的一种可重复编程的、通过不同的程序动作完成各种任务的高功能操作装置。现在的机器人大多是一种装有电脑的具有程序控制能力的多功能的操纵器，它们有的像人，有的并不像人。说它们是机器人，主要是因为它们具有人的手、眼、脑等器官的部分功能，把观测、思维和操作巧妙地结合起来，在减轻人的劳动强度、提高生产效率、改善劳动条件、改进产品质量、降低生产成本等方面发挥重大作用。另外，高温、高压、粉尘、有毒、放射等恶劣危险的生产环境下的手工操作，尤其是宇宙和海洋开发就更需要用类似人的操作功能的自动装置替代人从事各种操作。机器人具有较大的应用价值，对有关学科如控制论、信息

---

* 本文完成于1992年。

论、计算机科学、电子科学以及人工智能等的发展也有很大的促进和推动作用。因此，机器人受到人们越来越大的重视。现在，机器人是可编程的、多功能的机器，能在自动控制下，执行各种操作或移动任务，所以实际上机器人也是人的一种工具，主要是在功能上仿人的自动装置。

在机器人的开发中，综合和发展了机械工程、电气工程、计算机科学、控制理论、传感技术、人工智能、仿生学等有关学科，于 20 世纪 70 年代初形成了一门新兴交叉学科——机器人学。它是一门关于机器人及其设计、制造和应用的学科，是当今世界各国高科技竞争的热点之一。

但是，长期以来，机器人主要是作为玩具或戏剧、漫画、幻想小说的主人公。因此，人们常把机器人理解为虚构出来的过分像人的怪物。一提起机器人，总是有些令人生畏之感。就连最早提出友善的有用的机器人观念的美国幻想小说家阿西莫夫(Asimov)在他的幻想小说《我，机器人》里也提出了有名的机器人三条定律：① 机器人不能伤害人，或者看到人遇害而袖手旁观；② 机器人必须服从人给它的命令，除非这种命令与第一条戒律相抵触；③ 在不违背以上二戒律的前提下，机器人必须保护自身存在。

在机器人的研制中，美国动手最早。1954 年美国人德沃尔(Devol)发明了一种装置，称之为可编程序的关节型搬运装置，并申请了专利。1959 年德沃尔和英格尔保哥尔(Engelberger)进一步发展了这一概念，组建了 Unimation 公司，生产出来了世界上第一台工业机器人 Unimate(万能伙伴)，被广泛用在汽车装配线上。它与自动机床的不同之处在于它可以通过重复编程来完成不同的作业，可以被教会做某些工作，然后自动去完成。1963 年，美国机床制造公司生产出来了第一台商用机器人 Versatran(多用搬运)，它们实质上都是通用的操作自动机。与以往的专用自动机相比，机器人的突出优点在于具有柔性，即能再现人教给的各种作业，产品改变时只要改变程序即可。这两台机器人的问世，是机器人发展史上的一个里程碑。从此，机器人由科学幻想、戏剧和漫画中的主人公变成现实，并开始大踏步地跨入生产和科学技术领域。

1968 年，美国斯坦福人工智能实验室，集中力量开发了一台采用计算机控制的带有视觉传感器的机器人 Shakey。它能识别语言信息，能看到桌上放置的积木块，并能根据指令移动这些积木块。

以后，美国的机器人研制事业迅速发展，各种机器人系统相继问世，由工业机器人发展到智能机器人，到 80 年代，机器人的研制工作突飞猛进。除了经济原因外，主要是微型计算机的出现及其迅速发展，给机器人研制提供了坚实的技术基础。现在，智能机器人已从识别简单的积木发展到环境识别、规划决策，通过与人的语言对话来完成某种动作的阶段，并且已经从实验室走向实际应

用。

继美国之后，日本、前苏联和西欧也大力发展机器人。日本在 90 年代曾进行过一次有关机器人的“人口普查”，按当时的统计，全世界的机器人总数为 40 多万台，而日本就达 27 万多台，占机器人总数的 60%多。工业机器人的使用极大地提高了日本的劳动生产率，促进了日本经济的发展。

二　机器人的种类及组成

现在研制和生产的机器人种类繁多，分类方法也不一致。习惯上把机器人分为工业机器人和智能机器人两大类。前者的功能是执行程序所规定的操作，后者则能根据传感器信息认识环境，修改、细化或制订自己的行动规则以实现人指定的目标。

1. 工业机器人，又分一般工业机器人和遥控机器人两类。

(1) 一般工业机器人。一般工业机器人由手臂和存储装置组成，可按预先存储的程序重复地进行操作。现在一般工业机器人生产和使用得最多，技术比较成熟。这种机器人大部分在汽车制造、电子、机械、建筑等行业从事焊接、油漆、装配、包装、零件加工、搬运等专业性工作。以轿车车体点焊操作为例，每个车体的焊点 700 多个，对多品种小批量生产线来说，轿车的种类不同，焊点的部位和数量也不一样，过去靠人工操作，不但不能保证焊接质量，而且容易出错，用机器人进行自动点焊又快又好，而且用机器人比用人安全，费用少、产量高、质量好。此外，机器人还广泛用在机械制造、压力加工、铸造、锻造等方面。由于机器人能严格地按人规定的程序不休息地操作，对人“忠诚”，所以人们把机器人亲切地称为工人的“帮手”，生产的“能手”。

机器人的应用范围正在迅速扩大，除了汽车、机械、电机、塑料加工等行业外，煤矿用它来采煤，林业用它来伐木修树，商业用它来上货下货，银行用它来办理存取款，家庭和医院用它来照顾病人等。

(2) 遥控机器人。遥控机器人是指在恶劣或危险的条件下作业，由人在安全的地方进行远距离操纵的机器人，或由人操纵在海底和宇宙空间等等特殊场所进行作业的乘载型机器人。

在高温、高压、高湿度、易燃、易爆、辐射、有毒、水下、航天等危险和恶劣环境下工作，对劳动者的健康和生命构成严重的威胁。因此，如何把人从危险、恶劣的条件下替换出来，乃是自动化的重要课题，而机器人正是代替人在上述条件下进行各种操作的替身。

在放射性环境中，如在核电站里，机器人可用来检查、修复管道阀门等。在军事方面，机器人已用于侦察、布雷、排除爆炸物、装填弹药等。在建筑业中，已有一种爬壁机器人可用来修理墙面、擦洗窗户，机器人对于改善劳动条件、提高产品质量和劳动生产率、减少安全事故和人受危险环境的伤害等方面有明显的效果。

2. 智能机器人。60 年代，在工业机器人广泛应用和不断改进的基础上又产生了智能机器人，它是人工智能各种研究课题综合运用的试验场所。智能机器人具有感知和理解周围环境、使用语言（包括理解使用自然语言）和推理、规划和操纵工具的技能，并能通过学习适应环境，模仿人完成某种动作。70 年代，智能机器人已从识别简单三维物体阶段进入环境识别、规划决策，再通过自然语言与人对话来完成某种动作的阶段。智能机器人在某些特殊场合的实际应用正吸引着人们的广泛兴趣。

智能机器人由测量和识别环境的“眼”、根据环境情况作出决策的“脑”和进行各种操作的多关节机械“手”和“脚”组成，如智能汽车，其前面有两台摄像机，它们摄取的信息被送到一台道路图像识别装置里，从而能识别道路上相当范围内的路形、路标及障碍物，电脑能对各种可能发生的情况，作出适当的处理，求出最佳方案。试验结果表明，机器人驾驶汽车能在相当复杂的情况下自动快速行驶，这个驾驶汽车的机器人就是一种智能机器人。

智能机器人是智和力的工具，它的智能功能是由“人造脑”产生的。机器人的人造脑就是电子计算机（一般是微型计算机），其作用、功能及构成与人脑有很多相似之处，它是计算机控制的机器人的控制中枢，当操作人员发出工作指令时，经过人—机联系装置而加入到计算机中。同时，外界环境的状态信息和机器人的手和脚的行动信息通过感觉装置（在复杂情况下要再经过初步信息处理装置）加到计算机中，计算机可将这些信息存储起来，或者根据这些信息按照一定的性能指标准则来确定工作决策并控制操作，当感觉装置感到外界状态变化了，计算机重新修改控制规则来适应外界条件的变化。

机器人的发展过程中出现了三代机器人。第一代机器人以固定程序、示教再现及可编程序为主要特征。绝大多数工业机器人属于第一代机器人，如上下料、喷漆、搬运机器人等。第二代机器人带有视觉和触角，采用自适应控制，出现在 70 年代，典型的如计算机控制的装配机器人。具有人工智能的机器人称为第三代机器人，它们具有模式识别和作业规划能力，有丰富的感知能力，自己拥有知识库，可在复杂的环境中完成各种作业。目前这一代机器人正在大力开发中，如视觉控制的智能机器人手眼系统，是国家“七五”国防重点课题，它与只能从事机械重复动作的工具机器人相比，具有可根据环境条件变化改变工作程

序，以适应不同操作要求的特点。实现这个目标，必须使机器人有一双明亮的“眼睛”，来正确感知外界变化，并实现手眼协调动作。研制视觉控制的机器人手眼系统，是提高机器人的智能性和功能性的关键技术，也是当前自动化领域和信息处理领域的前沿性课题。据人民日报 1990 年 6 月 28 日报道，我国华东工学院化工计算机系科技人员于 1990 年首次研制成功视觉控制的机器人手眼系统，该系统采用了二维定性信息和三维定量信息相结合的识别方法，使摄像机在很大范围内任意移动时，都能快速、准确地识别形状变化很大的各种工作，其定位准确性的误差仅 1 毫米左右。专家鉴定认为，这个系统在透视因视点改变而形状畸形的扁平工件的识别和定位方面，具有国际先进水平。

三　机器人能超过人吗

以电脑为核心的机器人，由于它全面地模拟和延伸了人的劳动器官、感觉器官和思维器官，比起以往的劳动工具和认识工具，在功能上确实发生了质的飞跃。随着科学技术的进步，机器智能的发展将非常迅速，那么会不会有一天终于超过人的智能，并成为未来世界的主宰？这种顾虑看来是不必要的。因为机器人本身是人造的，机器人的工作程序是人编的，机器人在工作过程中是受人控制的，机器人是机器而不是人，它们没有生命，不可能完全脱离人而独立存在。人创造了机器人，同时也把自己的本领越来越多的交给了它，智能机器人的智能水平总是受着当时人类智能水平的制约。当机器人能够模拟人的某些思维活动的时候，人的智能又向前发展了，这个过程永远不会完结。

确实，机器智能在某些局部性能上是优于人类智能的，否则机器智能就失去了存在的价值。但是，它在全局上却永远不可能超过人类的智能水平，它不会也不可能具有比人更高级的创造能力，人不会消灭可以为人类服务的人工智能，也不会让人工智能来消灭自己。人和智能机器人可以平行发展，互相促进，但人永远是主导者。□

6 专家系统*

专家系统(expert system,简称 ES)是 20 世纪 70 年代后期知识工程的又一重大成果,也是人工智能领域中最活跃的分支学科之一。由于专家系统相对来说是一门较新的学科,人们对它的研究还很有限,但它的应用十分广泛,已渗透到许多领域,凡需要用专家知识来解决问题的地方,一般都可以用到专家系统。如专家系统应用于医疗诊断部门,就产生了医疗诊断专家系统,即计算机医生。因而专家系统越来越引起众多科学家的关注,特别是近年来,发展极为迅速,不但建立了许多实用的专家系统,还研制了专家系统开发工具和环境。

什么是专家系统?虽然它已有 20 多年的发展历史,但至今尚无一个精确的、全面的、众所公认的定义。粗略地说,专家系统就是把各行各业专家(一个或几个)的知识、经验和解决问题的逻辑思维功能,总结出来,形成一些有规则的东西,作为规则编成启发式程序,然后存入计算机,它们在计算机里可以模拟、代替或辅助专家的某些工作,解决实际问题。它的目的是使计算机在某些领域中起到人类专家的同等作用。所以,专家系统是一个具有大量专门知识和经验的计算机程序系统。

专家,一个人之所以被称为专家,其关键就是因为他掌握了关于某一领域的大量专门知识,在这些知识中,一部分是他从书本、资料或向他人学习来的,但主要还是他在长期实践中逐渐积累起来的经验性知识,才使得专家在

---

* 本文原载于《知识工程》1992 年第 4 期。

处理问题时比别人要技高一筹。

由此可见，如果计算机能够存储关于某一领域的一个或几个专家的专门知识，并能有效地利用这些知识去解决问题，那么，计算机也应该能很好地去解决该领域的复杂问题。专家系统的基本思想盖渊源于此。

其次，什么是启发式程序？启发的原意是“帮助发明或发现”，它们是一些策略、技巧、窍门或简化步骤，把这些总结成规则，编进程序就叫做启发式程序。对“启发”一词有不同的理解，广义地说，凡是有助于解决问题的办法，不管这些办法是否和人有关，都可以叫做启发；狭义地说，仅指和人有关的一些办法。这里是指后者，即考察人在解决问题时用些什么方法，有些什么策略、窍门或捷径，人们是怎样做的。根据人们的语言报告和行为的表现，总结出思维活动的规律，把这些规律编进程序，用计算机模拟，模拟的结果也可能和人的行为不一致，但经模拟、修改，再模拟再修改，就可以使计算机的工作和人的行为逐步接近。

1. 专家系统的特点

专家系统的特点是把专家解决问题的过程中使用的知识和经验，分成事实和规则，以适当的形式存入计算机，建立知识库，并采用合适的控制系统和推理规则，作出判断和决策，以专家的水平完成专门的、难度较大的任务。因为专家系统使用的工具是计算机，所以专家系统也可称为机器专家系统。

一个专家系统一般都是与使用者交互对话。它可以提出问题，并从使用者那里取得回答，所以又可称为专家咨询系统。这样的系统根据不同的输入产生不同的结果。

目前，世界上已研制出几百种人工智能的专家系统。如医疗诊断专家系统（计算机医生）、地质勘探专家系统（计算机地质勘探工程师）、教学专家系统（计算机教师）、秘书专家系统（计算机秘书）、军事指挥专家系统（计算机军事指挥家）等等。这些专家系统在处理某些特定类型的问题时，有的比人做得还好。如名医的经验是人类的宝贵财富，若将其加以整理，编成程序存入计算机，就构成医疗诊断专家系统，它会使名医发挥最大限度的作用。一个典型例子就是我国医务界同中国科学院合作研制成功的利用计算机进行中医辨证施治的“关幼波肝病诊断系统”（以下简称关系统），它意味着我国医务界向着医疗诊断自动化跨出了可喜的一步，这是一个了不起的成就。

关幼波是北京市中医医院著名的老大夫，对中医的辨证施治有很高造诣。他有 50 多年治疗肝病的经验，要把他的真本领学到手，不用多年的努力是难以奏效的。他根据多年的临床经验，将肝病分成 8 个主型，36 个亚型；各种脉象、脸色、胃口、舌苔、体温、出汗、化验的指标，总共有 180 多项；再考虑 150 多种中

药，按照中医的辩证施治，可以开出多达 19 亿个实用处方。让人来学，几年的时间是无论如何也培养不出来的。若用计算机来学，则轻而易举。首先由一定水平的中医和关幼波密切合作，用科学的方法将关幼波的经验加以整理、总结、描述出来，在此基础上建立数学模型；其次，就已建立的数学模型，研究出正确的算法，编成程序，并存储到计算机里。这样，一个专家医疗系统（即关系统）就建成了。有了关系统，只需要一名接受过中医基本训练的操作人员，就可以像关幼波本人那样为肝病患者进行诊断。当医生把病人症状、化验指标等输入计算机后，计算机就立即工作起来，仅 14 秒钟，对病人的诊断、处方、计价和编制病历档案等工作就全部完成了。经过对 266 名病人进行门诊试验，电子计算机的诊断和开出来的处方，经过和关幼波本人的治疗对比发现，符合率达 97.7%，得到了关幼波本人的肯定。

专家系统的一大优点是可以复制，因此，在边远的地方配上一个医疗专家系统，即使当地的医疗水平不高，但只要有能正确按脉问诊的大夫，就能达到像关幼波那样的水平。现在计算机接上几十或成百的终端是轻而易举的事，如在配有关系统的医院里，在每个终端配上一名实习大夫，就能够达到几十或成百的关幼波在同时问诊的效果。这样，一个专家系统的问世，就相当于一批具有同等作用的人类专家在不同地方，同时或不同时进行工作，为人类服务。因此，不仅促进了科学技术的发展，而且对人类社会诸方效益也有不可估量的作用。另外，专家的知识、经验过去主要是靠口传或书本记载；现在，可以把专家知识经验存储到计算机系统中去，即使专家去世了，他的知识和经验还可以照样为人类服务。因此，计算机专家系统为人类知识的积累和继承提供了新的方式。

2. 专家系统的组成

一个专家系统之所以能工作，首先是因为它有一个知识库和数据库。知识库是专家系统的第一个重要组成部分，用以存储从专家那里得到的关于某个领域的大量的知识和经验。数据库用于存储各领域的初始数据和推理过程中得到的各种中间信息。要完成组织专家知识、建立知识库和数据库工作，首先需要确定“知识表示”方法。“知识表示”是建立专家系统的关键环节，它与一个专家系统的建立成功与否密切相关。所谓知识表示，确切地说就是如何以计算机能够存储的形式来表示知识。另一个与此相关的问题，即所谓的知识获取问题，其重要性是不言而喻的，因为没有大量的专门知识也就无所谓知识库。知识获取就是把解决问题所用的专门知识从某些知识来源变换为计算机程序。可能的知识来源包括专家、书本、资料库和知识工程师自己的经验。目前，知识的获取主要是由知识工程师来完成的。知识工程是一个新概念，简单地说，就是利用知识的应用系统的有关技术，以知识为研究对象，其大部分工作都包含

着实践知识的获取和利用，即提取知识，然后以适当的计算机的形式表示它，再利用它。因此，知识工程师是一个计算机方面的工程师，他与某一领域的专家交谈，提出问题，记下专家的思路和解法，并把它以正确的形式存储到知识库里。获取知识是建立专家系统最困难的阶段之一，这一方面是因为专家系统是一个比较复杂的程序系统；另一方面，即更为重要的是，由于专家的专门知识在专家头脑中往往没有很好的组织结构，所以，要求专家在一个很短的时间内都整理、总结出来是很不现实的。一般是在总结一批主要的专门知识后，知识工程师先建立起系统的基本模型，再通过大量的实例，测试已获得的知识的可靠性，进而从中发现错误，并不断修正和扩充，直到系统逐渐完善。因此，建立专家系统的过程一般都是一个扩充性过程，是知识库不断健全和完善的过程。如果要真正达到专家那样高的水平，这一阶段要花费很长的时间。

专家系统的第二个重要组成部分是推理机。顾名思义，推理机是推理的机器，它具有进行推理的能力。从计算机科学来讲，推理机是一组程序，由它控制、协调整个系统，并根据当前输入的数据和知识库的知识，按一定的推理策略，解决当前的问题。即能根据知识推导出结论，给出建议，同时给出该结论的可信度，以供用户决策参考，而不是简单地给出现成的答案。其次，推理方法不可能完全独立于所要解决的问题的种类。但同时，推理过程又不能过于针对某一特定问题，以致其他场合不能适用。因此，我们所要寻求的推理方法既应该是有效的，可行的，同时又可普遍地应用于描述和解决广泛的问题中。专家系统中，通常所采用的推理机制有三种：正向推理、反向推理、正反向推理。正向推理是以原始数据为前提，按一定策略，运用知识库中的知识，逐步推出结论来；反向推理则从结论开始，先假设结论正确，再去寻找论据，看看结论是否真的正确；正反向推理是正向推理与反向推理的混合运用，它必须具备正向推理与反向推理的所有功能。

3. 专家系统的现状

20 世纪 70 年代后期，专家系统成熟，先后出现了一批卓有成效的专家系统，在医疗领域尤为突出，出现了多种多样的医疗专家系统。进入 80 年代，专家系统更有了突飞猛进的发展，大批新系统不断涌现。如美国已研制出一批“简易的”专家系统供在微机上运用，这是专家系统研制的突破。专家系统的应用领域也迅速扩大，甚至渗透到政治、法律、军事等重大决策部门。著名的美国兰德软件公司这些年来所研制的专家系统，也主要是集中在这样几个方面。同时所处理的问题的深度与广度也不断增加，正如人工智能专家费根鲍姆(Feigenbaum)所说：“80 年代是专家系统的黄金时代。”预计今后将有更多更好的专家系统产品问世。

我国专家系统的研究与开发工作起步较晚，大约始于70年代后期，但进展不慢。首先在医疗领域展开，随后很快进入农业领域，研制出小麦育种、水稻育种等专家系统。80年代初，我国专家系统相继渗透到交通运输、地质勘探、天气预报等领域，有交通运输调度、地质勘探、台风路径和暴雨预报等专家系统问世，有的已投入使用，并收到了良好的经济和社会效益。80年代中期，我国专家系统的研制与应用又进一步扩大到如数学、物理、化学、经济、法律、军事指挥等领域，先后在这些领域有一大批专家系统问世，出现了一种各行各业竞相研究与开发专家系统的局面。国家已经把实用专家系统、专家系统开发环境和知识工程列为"七五"攻关项目，每年都投入大量人力、物力开发各种专家系统，如清华大学开发出杂货船装载专家系统。杂货船的装载是个典型的数值处理问题。随着船舱体积的增大，装载货物量的增加，装载的方法也急剧增多，在装载的同时还要考虑到船只总体重量的平衡，船只的安全，货物本身的特性（如形状、大小、能否倒装、能否重压以及和其他货物的相容性等），使杂货船的装载成为一个多约束的复杂问题。该专家系统采用高效运行的、专门研究人工智能的LISP语言编程，同分舱、理舱、计算等六个子系统构成，其知识库有200多条产生式规则，较好地优化了杂货船的装载问题。

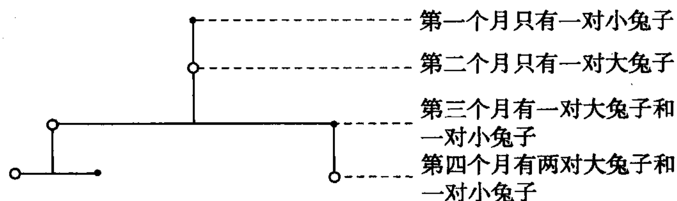
推理语言是一种通用专家系统描述语言，用它可以方便、快速地建造专家系统。目前在国际上已有数千个实用的专家系统，它们在众多复杂的、需要专家才能解决的领域中发挥重要作用。1991年9月，由中国科学院数学所计算机室陆汝钤等人研制成功的"通用逻辑推理语言及其实现系统"已通过专家鉴定，这是我国自行设计，并已达到实用水平的新型人工智能语言系统。这种推理语言还有效地解决了专家系统建造周期长、费用高的问题，好懂易学，即使没有计算机基础，也能很快掌握推理语言，快速建造自己的专家系统。推理语言系统已成为一个真正实用的高科技产品。

有人估计，在20世纪末，将会开发出一种崭新的专家系统，如除了基于成千上万条规则的专家系统外，还会有不规则化的专家系统，同时专家系统中的解释系统功能也会有很大的改进，不仅能解释推理，而且能够解释专家系统的行为和过程。□

新兴分支学科：迈向不断演进的独立命题

1 递归函数论和可计算性*

1228年，意大利数学家斐波那契（Leonardo Fibonacci）在其所著《算盘书》的修订本中提出一个有趣的"兔子生兔子"的数学问题：设想把一对兔子（一雌一雄，下同）放在一个地方，四面用墙围住，如果一对小兔子经过一个月就能长成大兔子，而一对大兔子经过一个月就可生出一对小兔子，问从刚出生的一对兔子开始，满一年时总共可繁殖成多少对兔子？这是一个算术问题，但却不能用普通的算术公式来进行计算。为此，我们用"○"表示一对大兔子，用"•"表示一对小兔子。在第一个月只有一对小兔子；第二个月有一对大兔子；第三个月这对大兔子生一对小兔子，加上原来那对大兔子，因此一共就有两对了；第四个月，那对小兔子未生育，而原来那对大兔子生下一对小兔子，因此这个月就有三对了。以上过程可用下图表示：



按此继续计算下去，就得知从第一个月到第十二个月的兔子的对数分别为

$$1,1,2,3,5,8,13,21,34,55,89,144。\quad (1)$$

答案是：满一年时总共可以繁殖成144对兔子。实际

---

* 本文原载于《自动化博览》1991年第3期。

上，兔子不会以这样快的速度繁殖，所以这只是一个假想问题。

由兔子问题所引出来的那一串数字(1)有一个明显的特点：从第三项开始，任何一个数都是其前面两个数的和。用 F_n 表示通项，则第三项 $F_3=2=1+1$，第四项 $F_4=3=1+2$……若令 $F_1=F_2=1$，而当 $n\geqslant 3$ 时，令 $F_n=F_{n-1}+F_{n-2}$，则得数列

$$F_1,F_2,F_3,\cdots,F_{n-1},F_{n-2},F_n,\cdots$$

此数列称为斐波那契数列。仅凭观察(1)想得到该数列的通项十分困难，但斐波那契还是推导出来了，即

$$F_n=\frac{\sqrt{5}}{5}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right]。$$

公式中虽含有$\sqrt{5}$，但用任何自然数(非零)代换公式中的 n，都能得到正整数的结果。例如取 $n=12$，代入后经计算即得 $F_{12}=144$。所以 F_n 实际上是以自然数为自变量的一类正整数函数，称为斐波那契函数。函数论中有广泛的递归函数与之很有相似之处。那么，什么是递归函数呢？

递归的来源较久，一个研究的对象直接或间接的调用自身的定义，则称为是递归的。上面所举的兔子繁殖问题就是利用递归概念的一个例子。假设第 n 个月共有 F_n 对兔子，其中新生的小兔子为 S_n 对，则有下面的关系

$$F_n=F_{n-1}+S_n,S_n=F_{n-2}。$$

所以

$$\begin{cases}F_1=F_2=1,\\F_n=F_{n-1}+F_{n-2},n\geqslant 3。\end{cases}$$

上面式子联合起来称为递归式。由递归式并利用前两项初值，数列$\{F_n\}$便可以完全确定下来。但直到 1931 年，由曾任美国普林斯顿高等研究所教授的哥德尔(Kurt Gödel)提出了著名的不完全性定理。为了给出这个定理的证明，才使用递归函数这一概念，并第一次给出了严格的而实际上是原始递归函数的定义。他当时没有区别原始递归和一般递归，只用“递归”这个词，可是，这种类型的函数不能包括一切递归函数。递归函数是一种数论函数，就是说，这种函数是取 0,1,2,…这些整数值的函数。它本来是数学中使用的一种函数。例如，在 1923 年，斯柯灵就利用它研究过初等数论；后来，又由哥德尔用于数学基础的研究，由此逐渐发展成一门独立的数学分支——递归函数论。半个世纪来，它的主要内容包括原始递归函数、一般递归函数、部分递归函数、递归可枚举性、判定问题、递归不可解理论、α 递归论、谱系理论等。

在数学中常用到一些递归式，借助这些递归式，一些复杂的运算可以还原为一些简单的运算，从而求出结果。递归函数的概念就是在此基础上形成和发

展起来的。由斐波那契函数我们可得

$$\begin{cases} F(1)=1, \\ F(2)=1, \\ F(n^+)=F(n)+F(n-1),\text{当 } n\geqslant 2 \text{ 时}, \end{cases}$$

其中 n^+ 表示 n 的后继数。例如取 $n=3$，则其后继数为 $n^+=4$，代入第三式，得

$$F(4)=F(3)+F(2)。$$

这样，$F(4)$就简化为 $F(3)$与 $F(2)$的计算，又利用(1)与(2)式得

$$\begin{aligned} F(4)&=F(3)+F(2) \\ &=(F(2)+F(1))+F(2) \\ &=(1+1)+1 \\ &=3。\end{aligned}$$

同法可求得 $F(5)=5$，$F(6)=8$，…由此可知，任给一自然数 $n(n\geqslant 1)$，总可以用上列的递归式一步一步地求出 $F(n)$的值。如果把这类递归式理解为一种算法，那么，我们所要求的结果，就能用这种算法一步一步地机械地求出来，这样一种解决问题的纯机械性过程，也称为能行性过程。因此，递归函数论有时也称为能行性理论，或能行性可计算理论。能行性可计算是一个直观概念，可以用普通语言解释或者举例使它较明确。我们说，能行可计算是一种机械过程，或者说能行的方法是：根据事先给定的规则在有限步骤内可以完成。从这个意义上说，递归函数论也是一种算法理论，它具有机械性或能行性的特点。总之，递归函数论是关于可计算性和可判定性的科学，是数学也是计算机科学的分支理论，它与数理逻辑、控制论以及某些哲学问题也有关。所以递归函数论对于非数学工作者来说，也具有一定意义。

常见的数论函数，如加、减(算术减)、乘、除(算术除)、乘方、最大公约数、最小公倍数等都是原始递归函数。它的应用是广泛的。后来，德国数学家阿克曼(Wilhelm Ackermann)在 1928 年写的论文《希尔伯特构造实函数的方法》里曾举出一个能行可计算函数，并证明其函数值的增长较任何原始递归为快。这说明了存在着非原始的能行可计算函数。1934 年，哥德尔根据艾尔伯朗的建议，提出一般递归的概念，于是原始递归函数开始向一般递归函数演变，发展了算式演算并给出了一般递归的严格定义。目前已发展为数理逻辑一个分支的递归函数论，一切原始递归函数都是一般递归函数，反之不然。如阿克曼函数是一般递归函数而不是原始递归函数。原始递归函数和一般递归函数的严格定义都比较抽象，对数学和数理逻辑知识不够的人来说是难以接受的。但由上面例子可知，我们可以把递归函数直观地理解为可以通过某种“回归”，将未知归

结为已知的一种函数。

下面谈谈算法。算法是一个古老的数学概念。所谓算法，从直观上讲，就是对于某一类问题，我们能够找到一组规则，运用这组规则，可以机械地在有限步内得到结果。算法在计算机里就体现为程序问题。上述的递归函数及英国杰出的数学家图灵所提出的图灵机等则是给出了算法的精确数学定义，但并不是任意一类问题都有算法的。那么，有没有一个有效的方法能判定一类问题有无算法呢？这就是所谓判定问题，或者能行可判定问题。它与算法问题是密切相关、难以分开的。一类问题的能行性，就是指它的可判定性，而判定的关键是要确定一类问题有无算法。说得明确些，一类问题 A 是能行可判定的，是指存在着一定的算法，运用这种算法，就可以在有限步内有效地判定这类问题中所预先给定的命题是真还是伪。所以问题 A 的能行性问题，就是给 A 找算法的问题。例如 1931 年，哥德尔证明了即使是初等数论中的命题类也是不可判定的。图灵提出了图灵机以后，从另一个角度给出了可计算性的严格定义。在这种定义中，图灵给出的不是一种函数，而是一种指令表，以此来说明能行可计算性。就是说，这种理想计算机可以用来计算全部的可计算函数。图灵的工作在方法论上最显著的特点在于它实质上是抽象地分解人的计算动作，得到了若干简单的但在性质上又明显的是机械的操作，并证明能把它们组合起来执行任意复杂的机械动作。这样，一个算法在质上增加复杂性的问题就变为在量上增加为实际完成算法所需的存储量和时间问题，从而为计算机软件理论奠定了最重要的基础。

递归函数论、图灵机理论等所研究的均限于理论上的能行可计算性。随着计算机科学的发展，人们越来越认识到，理论上可计算并不等于在实际中是可计算的。这是因为计算机要受到时间的限制。如果某类问题是能行可计算的，但计算步骤是按幂指数函数甚至是增长更快的函数增长，即使计算机的运算速度很快，也是无法在短期内得出结果的。另外，运算时间越长，则费用越高。因此，20 世纪 60 年代以后，现实可计算的概念就被人们提了出来。研究一类问题是否在实际上可计算，进而形成了算法分析或算法复杂性理论，这是可计算性与计算机科学相互渗透的产物，是个年轻的分支学科。今天，它正处在朝气蓬勃发展的时期。在美国的基础科学发展战略中，算法复杂性是应用数学的重点发展方向之一，卡玛卡多项式就是为解决复杂性问题而提出的一种新的算法，它的有效性已于 1988 年 9 月在东京举行的第 13 届国际数学规划讨论会上得到了公认。□

2 突变理论*

一 突变理论的由来

客观世界存在着两种基本的变化方式，一种是连续的变化，对于这种变化，在数学上早已成功地运用微积分的方法获得了圆满的解决；另一种变化是不连续的飞跃，如岩石的断裂、桥梁的倒塌、火山的爆发、水的沸腾等等，这是一些事物从性状的一种形式突然地跳跃到根本不同的另一种形式的不连续变化，我们称之为突变(catastrophe，也有人译为灾变)。不但自然界存在着许多突变现象，即使在生物界和社会科学领域内也有很多突变现象。例如，一只受惊的狗，因涉及到被逼迫的刺激而突然地放弃逃走的念头转为进攻(即所谓狗急跳墙)；市场上稳定的物价，因受到许多因素的影响而突然地出现急跌等。因突变问题而造成的不连续变化过程，传统的微分方程的方法是无力解决的。那么，有没有用来描述和解决各种飞跃和不连续过程的数学理论呢？这早就引起了人们的关注和研究。法国数学家雷内·托姆(R. Thom)在总结前人研究的基础上，于1972年发表了《结构的稳定性与形态发生学》的名著，运用微分映射的奇点理论来研究自然界中的各种形态、结构的非连续性突变，奠定了突变理论的基础，标志着突变理论的正式诞生。以后十多年来，经过英国著名数学

*本文原载于《数学通报》1986年第8期。

家齐曼(E. C. Zeeman)等人的努力,逐渐使突变理论成为数学中最年轻的分支之一,引起了国际数学家、哲学家、生物学家和社会学家的广泛关注。英国《大不列颠百科全书》誉其为“牛顿、莱布尼茨以来数学界的又一次智力革命”。因为牛顿用他的理论解释了所有的连续的、渐变的现象,而托姆的理论解释了所有的不连续的、突变的现象。

二 突变理论的主要内容

1980年,英国数学家桑德斯(P. T. Sounders)出版了《突变理论入门》一书,该书开头就指出:“作为数学的一部分,突变理论是关于奇点的理论。”奇点是相对于正则点而言的。一般地,正则点是大量的,而奇点则是个别的。但是,正因为奇点奇特个别,因而在数学中占有突出地位。在数学中,我们经常遇到求函数的极值问题,函数达到极大或极小值的点就是函数的导数等于零的点,这就是最简单的奇点,或称为临界点。三维欧氏空间中,临界点的集合是一个曲面,称为临界曲面。使函数取极小值的点叫稳定点,临界点不一定是稳定点,所以,临界曲面上的点可能使系统稳定或不稳定。事实上,利用数学方法研究突变现象由来已久,早在19世纪末法国杰出的大数学家庞加莱在非线性常微分方程的定性理论中就应用拓扑学方法研究过奇点。庞加莱与托姆都是法国的拓扑学家,他们的拓扑思想原是一脉相承的。托姆的突变理论所引用的例子,大都是庞加莱当年所处理过的。托姆的主要贡献在于以拓扑学、奇点理论为工具,说明了有的事物不变,有的渐变,有的则是突变,从而为突变现象建立了一系列新的数学模型,用以解释自然界和社会现象中不连续的突变过程,使庞加莱原来提出的问题得到了更新的、更全面的解释。因而,人们认为托姆的突变理论是数学上颇为新颖的成果。

在突变理论里,把描述状态的那些量称为状态变量,而把引起突变原因的连续变化的量称为控制变量。例如在气—液相变模型中,温度和压强是引起水突变原因的连续变量,故为控制变量;水的密度则为状态变量,密度高的状态对应着液态,密度低的状态对应着气态。

设给出势函数 V,我们用方程

$$\nabla_x V=0$$

定义其平衡曲面 M,其中下标 x 表示梯度。仅就状态变量 x 而言,这个曲面由 V 的全部临界点构成,即由系统的全部平衡点(稳定的或其他的)构成。我们又定义奇点集 S 是由 V 的全部退化临界点组成的 M 的一个子集。

假定系统动力学可由一个光滑的势函数导出，托姆用拓扑的方法证明：可能出现不同性质的不连续构造的数目并不取决于状态变量的数目（这可能很大），而取决于控制变量的数目（这可能很小），特别是变化在三维空间和一维时间的四个因子控制下的初等突变。概括起来只有七种不同性质的类型，而且其中没有一种牵涉到两个以上的状态变量（对后者是指选择一组 n 个状态变量，使得其中与不连续性有关的不会多于两个）。这七个类型的初等突变，按其几何形状被称为折叠型、尖点型、燕尾型、蝴蝶型、双曲型、椭圆型和抛物型。列如下表：

突变类型	控制变量	状态变量	势　函　数
折叠型	1	1	x^3+ux
尖点型	2	1	x^4+ux^2+vx
燕尾型	3	1	$x^5+ux^3+vx^2+wx$
蝴蝶型	4	1	$x^6+tx^4+ux^3+vx^2+wx$
双曲型	3	2	$x^3+y^3+wxy+ux+vy$
椭圆型	3	2	$x^3-xy^2+w(x^2+y^2)+ux+vy$
抛物型	4	2	$y^4+x^2y+ux^2+vy^2+wx+ty$

在这七个类型中，最简单的是折叠型突变，它的势函数是

$$V(x)=x^3+ux,$$

其中 x 是状态变量，u 是控制变量，相空间是二维的。

当控制变量是两个时，最简单的突变模型都是尖点型。尖点型突变应用最广，临界面容易构成，且比较直观，故以尖点型突变为例，分析如下：

尖点突变，托姆又称为黎曼—雨果尼奥特突变，它的势函数是

$$V(x)=x^4+ux^2+vx,$$

状态变量是 x，控制变量则有两个，即 u,v。相空间是三维的，稳定平衡曲面 M 由方程

$$V'(x)=4x^3+2ux+v=0 \qquad (1)$$

给出，如图 1 所示。

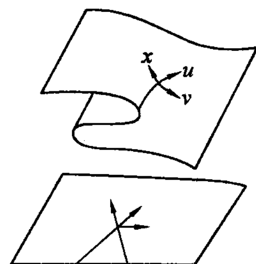


图 1

假若我们设想系统的状态是以 x,u,v 为坐标的三维空间中的一个点来代表的，则相点必定总是位于这个曲面上。事实上它必定总是位于顶叶或底叶。因为中叶对应于不稳定平衡。奇点集也是满足方程

$$V''(x)=12x^2+2u=0 \tag{2}$$

的 M 的子集。由(1)和(2)消去 x，得到

$$8u^3+27v^2=0,$$

从而找到分歧点集(所谓分歧点集乃是平衡曲面的折痕在 u—v 平面上的投影)。

例如气—液相变模型，在一个标准大气压下，水被加热到 100°C 即由液体转化为气体。这是一个突变，水的密度 ρ 是状态变量，控制变量有两个，即时间 T 和压强 P，故气—液的相变模型可用尖点突变来描述(图 2)。我们用 ρ,T,P 三个变量组成一个三维空间，温度 T 和压强 P 用两个水平轴表示，它们所确定的平面称为控制平面。与控制平面垂直的轴表示水的状态 ρ。水的密度变化可用一个特殊的曲面表示，即模型中具有折叠的那个曲面，称之为状态曲面。状态曲面上的每一个点表示一定的温度、压强下水的密度，折叠上叶代表液态，下叶代表气态，折叠区在底平面上的投影为一尖角型(此即尖点这一名称的由来)。我们设想温度与压强沿 AB 方向连续变化时，在相应的状态曲面上，水的密度沿曲线 $A'F'B'$ 变化。在起初阶段，水密度在折叠上叶连续下降，表示水密度在渐变，但到了折叠的边缘 F'，温度、压强只要稍顺着 AB 方向离开 F 点，水密度会突然跌到下叶的气态区域，发生不连续变化，这就是沸腾现象，它是一次突变；反之，如果温度与压强沿着 BA 的方向变化，可以看出，气态密度在状态曲面下叶沿 $B'F'$ 连续逐渐升高，但到了折叠的边界 F' 则稳定的连续曲面中断了，密度值突然上升到状态曲面上叶的液态区域，发生了不连续变化，这就是冷凝现象，也是一次突变。饶有趣味的是，从图 2 还可看出，如果温度、压强沿着 CD 曲线变化，绕过了尖角形的折叠区，相应的在状态曲面上的密度就沿着 $C'D'$ 平滑连续地变化，不出现飞跃而达到气态区。这就说明了为什么只要温度、压强绕过临界点，水的气化过程中，液态可以不经过沸腾而通过一系列中间过渡状态连续地变为气态。

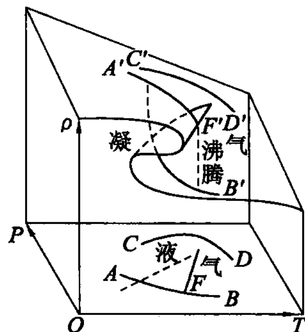


图 2

在描述狗的行为时，只要在状态曲面的上叶和下叶分别表示“攻击”和“逃跑”，控制变量是“恐惧”和“愤怒”，就可以用尖点型突变来描述“狗急跳墙”的突变。此外，冲击波的形成过程中的突变，弹性结构随着负荷越来越大而最后塌陷等，都属于受两个变量制约的，因此同样可以用尖点型突变来描述其变化过程。

尖点型突变还能说明为什么经济危机的爆发往往是突变，而危机的复苏又

是一种渐变过程。

齐曼还提出尖点型突变具有五条典型性质：第一是突跳，这在上面所举气—液的相变模型中已清楚地看到了；第二是滞后，如突变的原因消失之后，突变的结果仍会持续，要想做一次逆向的突变就需要比原有突变更强烈的动因；第三是发散，即突变现象是不可重复的；第四是双模态，即突变现象在同一原因的作用下可以有两个可能的行为；第五是不可达性。

其他的突变，如燕尾型、蝴蝶型、椭圆型、双曲型、抛物型等，它们都有三个以上的控制变量，一个或两个状态变量，它们的临界在三维空间中已无法划出，只能画出它们的截口或某些侧面。

三　突变理论的应用

借助突变理论不仅能加深对数学、力学、化学、工程技术等学科中已有定理或定律的认识，且已经获得了一些新的成果。特别是生物学、社会科学和人文科学中，许多现象很难用其他数学方法处理，却可以应用突变理论。所以它的应用范围是很广泛的。突变理论的数学基础是无可置疑的，人们对它的非议在于应用，主要是生物学、社会科学和人文科学至今还没有传统意义下的数学模型，不能对它们做定量的描述。下面举例说明。

(1) 在物理学中的应用。在物理学中借突变理论找到了光的焦散面的全部可能的形式，这是突变理论应用到光源研究中的著名的成果之一。当太阳光照射到一杯几乎是满的咖啡上时，设杯子的内部表面是单位圆，考虑单位圆中的任意点 $P(x,y)$，并设 $Q(\cos\theta,\sin\theta)$ 是这圆周上的任一点（见图 3）。

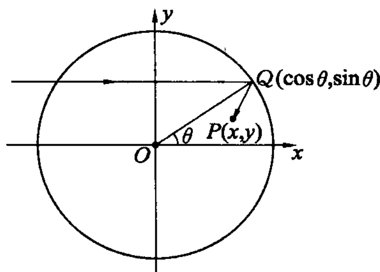


图 3

一般说来，在 Q 反射的光线将不通过 P，例外的是以下那些 Q 点（即那些 θ 值）。对于它们，由光源到 P 的路径长度有驻值，但不一定是极小值。如果光源

在 $x=-d$，又假若入射光线平行于 x 轴，则这一路径的长度是

$$g(x,y,\theta)=d+\cos\theta+\sqrt{(x-\cos\theta)^2+(y-\sin\theta)^2}。$$

驻值条件是$\frac{\partial g}{\partial \theta}=0$，于是反射光线族由

$$x\sin\theta-y\cos\theta=\sin\theta\sqrt{(x-\cos\theta)^2+(y-\sin\theta)^2}$$

给出，焦散线的方程可由方程$\frac{\partial g}{\partial \theta}=0$ 和$\frac{\partial^2 g}{\partial \theta}=0$ 消去参数 θ 而得到。这正好是以 θ 为状态变量，以 x 和 y 为控制变量时确定 g 的分歧点集的方程。因此，焦散线是以路径长度为势的系统的分歧点集。既然一共只有两个控制变量，用突变理论就可以证明这个突变是尖点型的。在这个问题中，势函数取极小值的意义是费马的“最少时间原则”，而突变集只是决定了焦散面的状态。

(2) 在哲学上的应用。一百多年前，德国哲学家黑格尔从大量的现象中第一次概括出从量变到质变的概念，揭示了由一种质态到另一种质态的规律。黑格尔认为量变是渐近性的运动，量变在一定限度内并不影响质，但是量变积累到一定程度时，就引起质的突然变化。质变表示渐近的运动的中断，使事物出现飞跃式的发展，而渐变和突变又受到一定条件的影响和控制。但是，一直没有出现阐述这种由量变到质变的数学理论，直到出现突变理论。突变理论就是从量的环节或侧面来反映和表现这种规律的。托姆的突变理论回答了为什么有的事物渐变，有的则是突变。我们知道，所谓稳定是指在微小的干扰下，事物不偏离稳定状态。用数学语言来描述，稳定性就是势函数趋于极小值。实践证明，任何事物的稳定状态必定处于某一“洼”的底部。所谓洼，在数学上就是势函数的极小值点。事物的渐变与突变过程可以看做“洼”的移动和消失(见图4)。



图 4

(3) 在社会科学中的应用。长期以来，社会科学被视为数学的禁区，但托姆的突变理论既可在自然科学，又可在社会科学中成功地应用。例如，涉及到狗的行为，是齐曼关于突变理论最早和最有名的例子之一。当你遇到一只陌生的狗的时候，这只狗可能逃走，可能无动于衷，也可能咬你一口。它的行为受到相互矛盾的两个因素控制：愤怒与恐惧。愤怒可由嘴张开的程度来衡量，恐惧则

由耳朵的耷拉程度来刻画。当恐惧压倒愤怒，它就逃走；当愤怒占优势，它就攻击咬人。但当愤怒和恐惧都很高时，就可能发生突变：一只本来似乎要咬人的狗突然掉头逃跑。同样，一只看上去将要逃走的狗会突然地转身咬你一口。数学家用几何学中的曲面概念提供了模型。在这个例子中，控制变量有两个，即愤怒和恐惧，状态变量是行为。此外，我们又发现上述狗的攻击行为具有尖点型突变的三个典型特征：双模态、发散和不可达性，故可用尖点型突变来拟合所观察到的现象。由于两个控制变量的效果互相冲突，坐标轴并不与正则尖点的 u，v 轴相重合，相反地，对狗的逼迫加紧时，光滑的响应会成为不可能。如图 5 所示，路径 1 和 2 表示单独的效应：逃跑或咬人，路径 3 或 4 表示突变的效应。

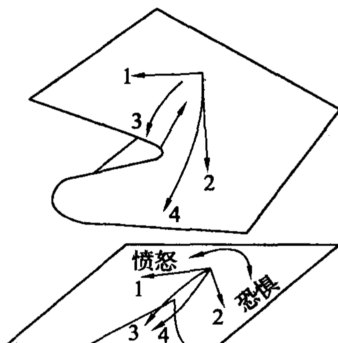


图 5

如果突变理论仅仅用于说明狗的行为，那就不值得称道了，令人惊奇不已的是，利用尖点型突变可以描述的社会现象还有很多。例如，与一个敌国冲突的一个国家政府，可以用“战争代价”与“威胁”来解释战争与和平之间的抉择；用“经济收益”与“人口密度”来解释某些城市的突然兴亡等，都可用尖点型突变来拟合。国外还有将突变理论用于描述股票市场的崩溃、大都市模型及城市发展模式的选择、蜜蜂的社会行为、细胞癌变等等。国内也有人用尖点型突变来说明中医的阴阳理论，通过治疗的实例，证实了“扶正祛邪”原则的正确性。

四　突变理论的争议和未来

托姆的突变理论发表以后，在国际数学界引起了激烈的争论，褒贬不一。褒者认为突变理论是继牛顿、莱布尼茨的微积分以后最大的发现，是数学中的一次革命。如《大英百科年鉴》1977 年版写道：“突变理论使人有了战胜愚昧无知的珍奇武器，获得了一种观察宇宙万物的深奥见解。”这就过分夸大了突变理论的应用。也有人加以抨击，认为突变理论所说的不连续现象可以用连续方法来解决，用微分方程研究激波就是一例。而对于社会科学中的突变现象又不能做定量描述。我们认为这都是偏颇之论。突变理论的数学基础是无懈可击的，争论的焦点是突变理论对社会现象还不能做定量的描述，这是事实。直到现在，还没有任何人能对社会的突变现象写出任何定量的描述，只是凭个人的体

验与感受到的一些特征，把突变现象归入到某类型的突变中去，这就难免主观臆测。突变理论是一种新的理论，能直接得到定性的结论，这就是一种突破。它当然不是具有这种性质的唯一理论，更不是处理不连续性的唯一理论。但突变理论应用之广泛，方法之奇特，是一般数学理论所无法比拟的。突变理论毕竟还是一门年轻的科学，还处在初创阶段，在理论上可能还不够成熟，在应用上存在这样与那样的不足与缺点，都是在所难免的，这也正是我们所应继续研究和克服的。我们相信，突变理论将在百家争鸣中不断充实、改善和发展。□

3 分数维几何学*

一　分数维几何学的缘起

中学里所学的几何学是两千多年前古希腊学者欧几里得(Euclid)所创立的，叫做欧氏几何学，通常又称为初等几何学。它有一条著名的平行公设，即第五公设。人们从生产实践中总结出来的这一公设，是不证自明的。但是否能用欧氏几何中的其他公理去证明它呢？许多学者试证都没有成功。1826年，俄罗斯数学家罗巴切夫斯基(Lobachevsky)在试证失败后断言：第五公设是一个独立的公理，不能由其他公理推导出来。于是他建立了一个在逻辑上既不矛盾但又不同于欧氏几何的完整的几何体系，后人称之为罗巴切夫斯基几何学。凡不是欧氏几何体系的几何学，统称为非欧几何学，它包括罗巴切夫斯基几何学、黎曼几何学。与之差不多同时发展起来的，还有解析几何学、微分几何学、射影几何学等。

随着科学技术的不断进步，人们在生产实践中对客观世界的认识越来越深刻，于是又产生了许多新的几何学，如代数几何学、算法几何学、整体几何学等。以上这些传统的几何学，总称为经典几何学。

几何学研究的对象是现实世界中物体的形状，如初等平面几何学主要是研究点、线；平面解析几何所研究的主

---

* 本文原载于《辽宁工学院学报》1989年第1期。

要对象是一次曲线(直线)与二次曲线;微分几何所研究的对象是光滑的曲线与曲面;代数几何学所研究的是复空间中的代数曲线。这些几何图形都是规则而光滑的,或至少是分段光滑的。但自然界的景象绚丽多彩,比如:天空中变幻无常的云朵,令人眼花缭乱的繁星,弯弯曲曲的海岸线、溪流,起伏不平的群山,滚滚的浓烟……尽管这些物体的本质千差万别,但有着一个共同的特点:很不规则,极不光滑。即使素以严谨和高度概括而著称的经典几何学,要描绘它们也有些无能为力。20 世纪 70 年代,获巴黎大学数学博士学位的法国数学家曼德尔布罗特(B. Mandelbrot)在法兰西学院讲课时提出了分数维几何的设想。后来他在美国哈佛大学任教时,受 IBM 公司之聘,在检查电子信息的导线时,发展了分数维几何学的概念。他发现所检查的许多问题中,存在着一种自相似性,即"部分与整体的形状相似",这使得他创立了分数维几何学。

分数维几何学,英文为"fractal geometry"。"fractal"一词源于拉丁文"fractus",与英文中的"fraction"(分数或碎片)及"fragment"(碎片)具有相同的词根。据曼德尔布罗特的解释,意为"不规则的"或"支离破碎的"。因此分数维几何学又被称为断裂几何学或不规则几何学。顾名思义,分数维几何学是以极不规则的几何图象为研究对象的一门几何学。在诸如大地形貌、晶体表面、晶界形貌、大气湍流、气象分布、催化剂结构、材料断裂机理分析、凝聚体结构等涉及物体表面几何形貌的研究中都有其重要的应用。饶有趣味的是,利用分数维几何学的原理编制成的复杂的计算机程序,可描绘出一些绚丽的大自然景象,简直可以与艺术大师的杰作媲美,真是令人赞叹不已。

二　分数维曲线

在经典的几何学中,点是零维的,直线或曲线都是一维的,平面是二维的,而立体则是三维的。这种维数,只取正整数,是拓扑意义下的维数,但不排除客观世界中维数为分数的物体的几何形状。的确,分数维(fracal demension)作为刻画几何形状复杂程度的量,在大多数情况下不是整数。先以雪花为例来说明分数维曲线的构造过程。

雪花在结晶过程中产生许多美丽的形状,但仔细观察它最有代表性的是六花片状结晶体,即六角形。我国早在西汉就有文字学家韩婴研究过雪花的形状,他说:"凡草木花多五出,雪花独六出。"1904 年,瑞典数学家科赫(H. von Koch)提出描述雪花的几何方法,所得曲线称为科赫雪花曲线。

取一个正三角形作为源多边形,取一条线段作为生成线,如图 1 所示。其

构造方法如下：

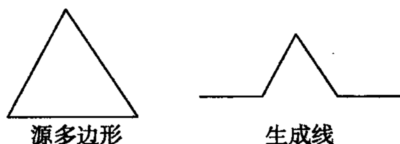


图 1

由图 2(a)开始，凡是源多边形的直线段（在这里就是正三角形的三条边），均按生成线的形状变形，变形一次就成图 2(b)，形成一个类似雪花的多刺状图形。再将这个图形中的每一条直线段按生成线的形状变形，遂成图 2(c)。如此变形下去。第三次变形生成图 2(d)，……随着变形的进行，图形边界曲线的长度可变得无穷大，但曲线所围面积却是有限的，即为原三角形面积的$\frac{8}{5}$。这种曲线的极限情形处处连续，但处处不可微。这就是著名的科赫雪花曲线。

可以看出，在上述构造过程中，每次变形实际上都是将生成线上的小三角形放在原来图形的外部。如果将小三角形放在原来图形的内部（这当然也可以），结果就得到了反雪花曲线，其长度也趋于无穷大，但曲线所围面积还是有限的，即为原三角形面积的$\frac{2}{5}$。

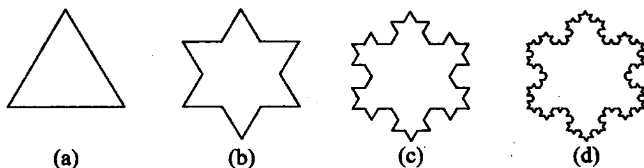


图 2

分数维几何学所研究的对象及方法正是这样。当然，也可以取正多边形为源多边形，而构造的规则是在每条直线的中间三分之一处拼接一个同形状的小正多边形，这也可得到一条形状极其复杂的分数维曲线。例如，在一个正方形的每条边中间向外拼接一个小正方形（边长为原正方形边长的$\frac{1}{3}$），这样一次一次地构造下去，就可以得到一个十字形的刺绣状图案。它的周长趋于无穷大，但曲线所围面积有限，即为原正方形面积的 2 倍。类似地，也可以向正方形的内部构造，结果就得到一个反十字形图案。

类似的构造也可以在空间进行。如将正四面体的每个面分成四个小正三角形，以中间的那个小正三角形为底，向外拼接一个小正四面体，这样继续构造下去，所得到的是一个多刺状立体。其表面积趋于无穷大，但内部体积仍然有

限，即为原正四面体体积的 3 倍。如果对立方体做这样的构造，结果得到一个凸刺状立体。

上面所说的源多边形应理解为广义的，但一条直线段也可以作为源多边形。图 3 就是以直线段为源多边形，以夹角为直角的折线段为生成线而进行构造的过程。变形一次成图 3(a)，变形两次而成图 3(b)，变形四次即成图 3(d)，变形八次遂成图3(e)。其极限曲线称为 G 曲线。

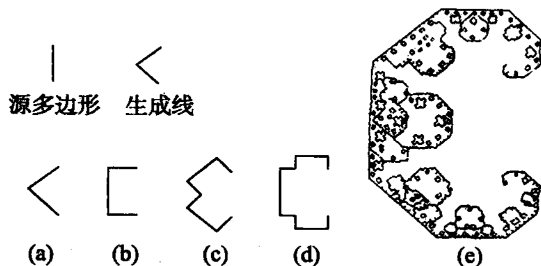


图 3

在经典几何中所研究的图形（曲线或曲面）是处处连续，至少是分段光滑，但以上所举的各种分数维曲线和曲面却与此不同，它们有一个共同特点，即处处连续，但无处光滑。

分数维曲线除去形状复杂这一特性外，还有一个很奇特的性质，叫非直观性质。如图 4(a)以一个正六边形为源多边形，以一条 Z 字形的折线段为生成线，折线段的每条线段的长为正六边形边长的$\frac{\sqrt{7}}{7}$。变形一次得到一个非凸的正十八边形，其面积与原图形相等；变形两次得到一个非凸的五十四边形。每次变形都使边数增加到三倍，而面积却保持不变。更使人惊奇不已的是，每次变形后的图形都可分割成七个与自己完全相似的小极限图形，如图 4(b)所示。因此，小极限图形与极限图形的面积之比为$\frac{\sqrt{7}}{7}$。

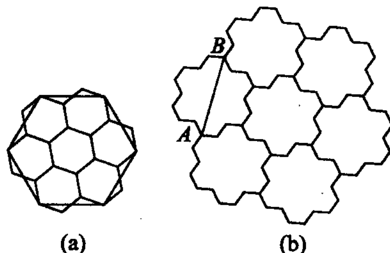


图 4

三　分数维的定义

拓扑意义下的维数不能反映分数维曲线的特征。如果把拓扑意义下的维数记为 D_r，那么上面所指出的各种分数维曲线的 D_r 都为 1。现在根据以上各种分数维曲线的特征给出分数维曲线的分数维数的定义。

定义　设一分数维曲线的生成线是一由 N 条等长的直线段连接而成的折线段，若生成线两端的距离与这些直线段的长度之比为 $\frac{1}{r}$，则定义这分数维曲线的维数为

$$D=\frac{\lg N}{\lg \frac{1}{r}}。$$

利用这个公式可以算出科赫雪花曲线的分数维为 $D=\frac{\lg 4}{\lg 3}\approx 1.2618$，反雪花曲线的分数维数为 $D=\frac{\lg 2}{\lg \sqrt{2}}=2$。图 4 所示曲线，其分数维数为 $D=\frac{\lg 3}{\lg \sqrt{7}}\approx 1.12915$。图 5 是曼德尔布罗特给出的一种雪花曲线，它的源多边形是正方形，生成线由 8 条直线段连接而成。除了当中 2 条直线段对接成一直线外，其余的相接成直角。图 5(a)及图 5(b)是分别按此生成线变形两次和三次后所得到的图形，其分数维数为 $D=\frac{\lg 8}{\lg 4}\approx 1.5$。由此可知，曼德尔布罗特雪花曲线的分数维数($D=1.5$)大于科赫雪花曲线的分数维数($D\approx 1.2618$)。由两图可见，随着变形次数的增加，曼德尔布罗特雪花曲线趋于复杂的速率比科赫雪花曲线快。因此，分数维数可作为描述分数维曲线复杂程度的一个参量。

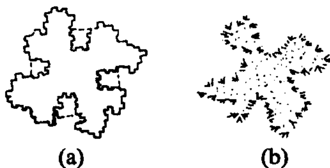


图 5

总之，以上这些分数维曲线的维数 D 都大于 1 而小于或等于 2，这是分数维曲线区别于普通曲线的一个重要特征。若在空间，还有大于 2 而小于或等于 3 的分数维曲面，大于 3 而小于或等于 4 的分数维立体。

四 应用

在自然界中，有规则的现象是特殊的、近似的，而无规则的现象却是普遍的。因此，分数维几何学所涉及的领域和用途是极为广泛的。它已在物理、化学、生物、地质矿产、医学、天体物理、哲学等领域中显示了强大的威力，受到了和正在受到不同学科的自然科学家和社会科学家的关注。

分数维几何学最奇妙的性质是“自相似性”。海岸线、云朵、树甚至描述经济现象的图形，乍看起来极不规则，但仔细观察，却都有很好的自相似性。这种图形的局部与整体的形状相似，局部的局部也与其整体的形状相似。例如，复杂的运动（指混沌），若用适当的坐标画出其轨迹，则复杂轨迹构成的复杂图形也有这种自相似性。悬浮液中某粒子的布朗运动轨迹、宇宙中的星系分布，都是一个分数维几何学问题。人类在对各种结晶进行观察和研究的过程中，逐渐形成了几何结晶学这一门实用科学。几何结晶学实为分数维几何学。曼德尔布罗特认为，自相似性为自然界的普遍规律。

在社会科学方面也不乏自相似性的思想。我国古老哲学思想中有“一粒米中藏世界，半边锅中煮乾坤”的谚语，此外，阴阳五行，五行生万物，五行的每行又有五行等，无不反映了这种自相似性的思想。1986 年，有人提出人体的结构与功能、蛋白质等也有这种自相似性，即分数维特性。在物理和天体物理方面，时空也可能是分数维的。诸如此类，不胜枚举。

总之，分数维几何学是一门具有强大生命力和广泛应用前景的新学科。虽然一创立它就得到一些具体应用，但毕竟是一门年轻的学科，存在缺点和问题也在所难免，这恰是后来人所应努力充实和开拓的。笔者相信，随着人们对自然现象认识的逐步加深，分数维几何学必将产生更加深远的影响。□

4 协同学*

协同学（Synergetics）一词来源于希腊文，它的含义是协同作用的科学。联邦德国理论物理学家哈肯（H. Haken）在研究激光理论的基础上，于 1977 年出版了《协同学导论》一书，标志着一门新的综合性横断科学——协同学的诞生；1983 年又出版了《高等协同论》，使协同学进一步完善，其学科理论称为协同论。这两本书是协同学奠基之作，刚一出版便得到世界范围的承认，被译成多种文字出版，成为风靡一时的畅销书。协同论与普里高津的“耗散结构论”、托姆的“突变论”构成区别于以前的“系统论”、“控制论”和“信息论”的新三论。1981 年，美国富兰克林研究院鉴于哈肯在协同学方面的开创性工作，授予他迈克尔迩奖章。其后，联邦德国总统又授予他联邦大十字勋章，以表彰他在科学上的卓越贡献。

协同学以现代最先进的理论（信息论、控制论、突变论）为基础，同时采用了统计学和动力学考查相结合的方法，通过类比，研究一个与外界有物质、能量、信息交换的非平衡开放系统（物理、化学、生物的，或者社会、经济的）。由于其内部子系统之间的相互作用，一旦外界控制参量的变化达到一定的阈值时，通过子系统之间产生的协调作用和相干效应，不仅可以从无序混乱状态变为宏观有序状态，而且也可以从有序变到混乱。这里所说的“混乱”，指的是决定性方程所描述的不规则运动。一个非平衡的开放系统，当外界参量逐渐增大到一定程度时，便出现了混

* 本文原载于《自动化博览》1990 年第 1 期。

乱现象。

任何一个远离平衡的开放系统，只要它由许多个子系统所组成，则在一定条件下，都可以从无序变到有序。但近几年来，人们还发现即使是远离平衡的开放系统，在一定条件下也可以从有序变到混乱，变到无序。最显著的例子就是流体力学中的湍流现象。在流体力学中，当流体绕过一个圆柱体流动时，在低速下流动是层流，在高速下突然出现了一个新的静态花样。随着流速的不断增高，出现了动态花样，此时涡流在振荡。最后，在更高的速度下，就出现了不规则的花样——混乱。著名科学家钱学森写道："哈肯用统计力学的方法来解决复杂系统的有序化问题，他严格证明，在一定条件下，这个有序化的出现是不可避免的。""在非线性的系统里头，还有这么一个可能，就是从有序也会变成无序，变成杂乱。""假设流体慢慢地流过一个物体，那么这流速是有序的；假如流速增加到了一个数值，稳定的、平衡的流动不可能继续下去，就要发生紊乱的流动，叫紊流或者叫湍流。"即出现了无序。

在协同学发展过程中，激光起着关键作用，哈肯形象地比喻激光理论是协同学的"路标"。激光是由激光器产生的，原意是因受激发射而产生的光放大。激光系统是典型的非平衡开放系统。一个典型的激光器主要由晶体棒和充满气体的玻璃管组成。组成激光材料的原子受到外界激发，俗称"泵浦"，便会辐射光波。当泵浦供给能量很小时，激光器就似一盏普通的灯，发射一些完全不相干的光波；但当泵浦输入功率超过某一阈值时，它就会发射出相干波列。这表明激光器的内部状态完全改变了，原来是无规则振荡的原子发射系统，现在以完全自组织的方式发生相同振荡。

在生物系统中也有类似现象。比如所谓"生物钟"这种有节奏的韵律现象也在许多生命过程中显现出来。在生物活体中，通过代谢给它供给能量。当所供给的能量超过某阈值时，便出现了有序的相干振荡，即出现了长程的位相相关，而这正是生物活体中具有极其惊人的有序性和极高效率的例证。最新研究结果表明，不论是宇观系统、宏观系统或是微观系统，只要它们是开放系统，则在一定条件下，都呈现出非平衡的有序结构，从而都可成为协同学的研究对象。

总之，协同学的中心课题，就是在生命和非生命这类开放系统中，不仅研究从无序到有序的演化规律，而且也研究从有序到混乱的演化规律。哈肯第一次真正将有序与无序统一起来了，确是独树一帜，具有鲜明的特色。哈肯在此基础上，建立了一整套数学模型和处理方案，从而可把一门学科中所取得的成就很快推广到其他学科中类似现象中去，这就是类比。

一个复杂的系统，其状态变量可能成千上万，影响事物发展的因素也多不胜数。如何描述系统的无序和有序？或者说如何描述事物在临界点附近的行

为？对此哈肯具体分析了系统中不同状态变量在临界点处的情况，发现状态变量分为两大类：一类是大多数在临界点附近阻尼大，衰减快，对相变的整个进程没有明显的影响，这类变量叫快变量或快弛豫参量；另一类不仅不衰减而且始终左右着演化过程，这类变量称为慢变量或序参量。序参量是指描述一个系统有序程度的量。在不同的系统中，序参量代表着不同的物理意义，比如在铁磁体（如罗盘的磁针）中，总的磁化强度可作为序参量。当自旋方向是无序时，序参量是零；而当所有的自旋方向是完全有序时，序参量达到最大值。在流体力学中，序参量表示速度和密度。在企业管理中，序参量可以代表工资、利润等。在一般情况下，描述系统有序状态的序参量只有很少几个，有时甚至只有一个（如单模激光），但它们始终主宰着系统演化的全过程，决定着演化结果出现的结构和功能以及系统的相变特点。这对所有系统无一例外。因此，如果要描述这类系统的相变，只需抓住序参量就行了。

为了正确地找出描述系统有序状态的序参量，首先要找出系统的各个状态变量，然后在这些状态变量中找出哪些是变化快的量，再利用协同学的原理——支配原理消去快变量，留下慢变量作为序参量。整个系统的性质用一个或几个序参量来描述，也就是说，在整个系统中，用很少的自由度来代替子系统中很多的自由度。这种利用状态变量随时间变化快慢的办法来区别序参量的做法，正是协同学独到之处。因为世界上任何一个有限大小的具体系统都有产生、发展与消亡的演化史，即都有一定的寿命。不仅如此，寿命在一定条件下还可以通过外参量来改变。所以说这种利用寿命的量作为序参量的绝妙方法，确实是独具特色的。

哈肯又利用严格的数学推理分析了当控制参量使系统偏离平衡态较少、系统内部的变化趋势线性地依赖于产生变化的作用力时，系统不会发生结构或功能的变化；当控制参量变化到一定阈值时，系统性质发生变化，系统内部的相互作用会形成新的有序结构；当控制参量继续变化时，有可能再形成新的结构，最终甚至会形成混沌而丧失一定的功能和结构。控制参量使系统处在远离平衡态较近的线性区，虽然系统也会受到各种扰动，但扰动只能使系统回到原来的均匀状态。控制参量超过临界值太远，也会使系统趋向混沌，仍然形成不了新的有序结构，只有当系统被控制在临界点附近，它才能通过内部的协同作用而达到有序。

子系统的复杂运动可分成两类：一类是子系统独立的杂乱运动，另一类是有序的彼此互相关联的运动，这两类运动构成了整个系统的宏观相变运动。当独立运动占主导地位时，系统处于无序状态。一旦外界控制参量达到某一临界值，子系统之间的协调作用、自我组织、系统的关联运动就占主导地位，系统这

时进入有序状态。如水的热运动:当独立运动占主导地位时,系统处于无序状态,如水变成汽;当关联运动占主导地位时,系统就进入了有序状态,如水变成冰。

协同学的内容十分丰富,还是一门正在发展中的学科。它利用现代数学中的突变理论反映系统内部协调的具体机制,并根据实际进行协调,使系统达到最优。由于其研究方法充满了辩证法思想,故对社会的协调持续发展,也具有指导和借鉴意义。□

5 混沌理论及其应用*

混沌是近年来数学、理论物理学、系统动力学等不同学科的学者们开始研究的一种新的运动形态。因为对它研究的历史还很短，对混沌迄今尚无一个确切的定义。协同论创始人哈肯(Haken)曾指出："混沌来源于决定性方程的无规则运动。"我国著名科学家钱学森称："混沌是宏观无序、微观有序的现象。"一般地，混沌是指一个系统所处的一种无序的、非线性的、不可测的状态。研究、分析、解释这种状态的理论，总称为混沌理论或混沌学。

我国有"渺渺蒙蒙不分上下，昏昏沉沉不辨内外"的谚语，其实说的就是混沌。混沌学是一门对杂乱无章的整体现象进行研究的学科。在过去只能看到随机性、稳定性和周期性的地方，这门学科提供了一种发现秩序和结构的新方法，它在一类"紊乱"现象的背后出人意料地发现了惊人的规律性。实际上，事物在空间和时间中的汇集方式，无不暗示着某种规律性，并大都可以某种数学方式(例如分数维几何学)表示它们的特性。混沌学与相对论、量子力学同为20世纪的三大科学发现之一，这不能不引起科学家们的广泛关注。但混沌学对物理学家来讲，是一门过程的科学而不是状态的科学，是一门演化的科学而不是存在的科学。

在自然界及社会经济领域内，混沌现象无时不有，无处不在。例如，激光是一种典型的非平衡开放系统，气体激光器在泵浦功率达到一定阈值时，会发出激光束；当泵

* 本文完成于1994年。

浦功率超过第二个阈值时，激光器又会发出很强且短的周期脉冲。在不同的条件下，光辐射又会变成混沌或湍流，形成滚筒状的所谓“具纳德花纹”。在继续加温时，振荡运动会失去完好的周期性，呈现出一种表现无序的湍流。此外，徐徐上升的青烟，变幻莫测的股市，复杂多变的气候，千姿百态的云彩，汹涌澎湃的洪水，在它们的背后，都有一些普遍适用的规律。看似混乱一片，实为秩序井然，这就是混沌。正因为无序是现象，有序才是本质，所以“天有不测风云”变成了天气可以预测；在令人眼花缭乱的股市风云中也必然有着秩序和规律。混沌学的问世也将反映一个整体的、更加具体而真实的世界，并可标志人类历史上又一次重大的科学进步，人们可以从更接近实际的角度去认识世界——这个既无序而又有序的世界。

其实，对无序现象的研究可以追溯到 19 世纪。法国数学家拉普拉斯建立了几率论；俄国数学家柯瓦列夫斯卡娅在用测量小偏差增长率的平均值来定义动力不稳定时，已向混沌的独立理论迈出了一步。另外，许多混沌理论的现代研究都能追溯到 19 世纪末法国数学家庞加莱在天体力学方面所做的贡献。然而比较有系统地对混沌现象进行研究的是耗散结构理论和协同论。

耗散结构理论是比利时物理化学家普里高津(Prigogine)于 1969 年提出的一种关于非平衡系统的理论，该理论指出：一个远离平衡的开放系统(无论是物理的、化学的、力学的、生物的乃至社会的、经济的系统)通过不断与外界进行物质与能量的交换，在外界条件的变化达到一定程度、系统的某个参量变化达到一定阈值时，通过涨落发生突变即非平衡相变，就可能从原来的混沌无序状态，转变到一种在时间上、空间上或功能上的有序状态。这种在远离平衡的非线性区形成的新的有序结构，由于需要不断地与外界发生物质与能量的交换，才能维持它的有序状态，故而耗散物质与能量，因此，普里高津等人称之为耗散结构。这种耗散结构系统，通过内部相互作用，自行产生组织性与相干性，称作自组织现象，所以这一理论也被称为非平衡系统的自组织理论。普利高津认为，一个系统要从混沌转变为有序，至少需要四个条件：① 系统必须是开放的，也就是说必须能从外界吸收负熵；② 系统必须远离平衡态；③ 系统的不同要素之间必须存在着非线性的相互作用；④ 系统必须通过涨落才能达到有序。普里高津认为，混沌是系统达到更高层次有序的源泉，有序和组织可以通过系统的“自组织”过程“自发地”从无序和混沌中产生出来。

协同学是德国著名理论物理学家哈肯于 20 世纪 70 年代初创立的一种非平衡系统的自组织理论，它的核心思想是协同导致有序，中心课题就是在生命和非生命这类开放系统中，不仅研究从无序到有序的演化规律，而且也研究从有序到混沌无序的演化规律。哈肯认为，由于系统内部各要素之间的协同作

用，在外部控制参量达到一定阈值时，系统可由原来混沌无序的状态转变为宏观有序的状态。哈肯第一次将有序与无序用辩证的观点将它们统一起来了。70年代后期，人们通过进一步观察与研究，认识到系统内各步的运动是无序的，系统的每下一步运动是不可预测的，外部因素对系统的影响也是非线性的。在某一特定的时刻，一个细微的外部影响将导致系统的巨大变化，这种系统又是封闭的，它与外界进行着物质与能量的交换，随机性与偶然性在这种系统的运动中起着很大的作用。一个著名的例子可较形象地说明这个问题：一个受到下面两个同样磁铁吸引的摆锤慢慢地运动接近两个磁铁的中点时，每个磁铁对它的吸引力差不多相等，那么它下一步对它现在的位置和速度的细微变化就极敏感了，于是运动就变得混沌而不可测了。一个小的变化导致以后更大的改变，这就是混沌的特点。对这种系统的运动，牛顿早期的经典理论是无能为力的。这就促使人们去寻求一种新的理论，于是各种研究和分析混沌现象的理论即混沌理论就应运而生并发展起来了。

于是，1979年哈肯等人通过观察与研究，认识到混乱现象的重要性，并发现一个非平衡(严格说是远离平衡)的开放系统不仅可以从无序到有序，而且系统还会走向自组织的高级阶段，即从有序走向混沌。值得注意的是，这里所说的混沌是指由决定性方程所描述的不规则运动，其中最显著的一个例子就是流体力学中的湍流现象。什么是湍流？它是在所有标度上的一种无秩序的混乱，大旋涡里套着小旋涡。它是不稳定的，是高耗散性的，这就意味着湍流消耗能量和产生阻力。它的运动是随机性的。在流体力学中，当流体绕过一个圆柱流动时，在低速下，流动是层流；在高速下，突然出现了一个新的静态花样。随着流速的不断增高，出现了动态花样，此时旋涡在振荡。最后，在更高的速度下，就出现了湍流，即出现了不规则的花样——混沌。

一个非平衡的开放系统，当外部参量增大到一定程度时，便出现混沌现象，这是协同学的新进展。事实上，混沌的研究打破了学科之间的藩篱，更新了观念，给人类整个知识体系一个巨大的冲击。混沌开创了非线性这一新学科。非线性科学强调普适性，强调局部和整体的统一。非线性科学的基本概念和方法已渗入到物理学、化学、生物学、地学乃至社会科学领域。混沌不仅成为一门理论，而且还成为一种方法，一种从事科学研究的方式。

混沌现象目前已被更多的人所认识，许多科学家也把混沌理论运用到自己的研究中去。如物理学家用混沌理论来指导他们对原子内的电子运动、分子和气体内的原子运动进行研究；生物学家用混沌理论来研究昆虫和鸟群的变化、传染病的传播、细胞的新陈代谢等；天文学家用混沌理论模拟早期宇宙的脉动，模拟星系中恒星的运动，以及模拟太阳系中行星、彗星的运动；工程设计人员在

设计中也考虑混沌现象的干扰等。由于电子计算机的使用,混沌理论在数学领域内更可望一显身手。

混沌理论的创立,把系统表现的随机性和系统内在的决定性机制巧妙地结合起来。混沌的现代研究使人们渐渐明白,十分简单的数学方程完全可以模拟系统如同瀑布一样激烈的行为。系统的输入端微小的差别能够迅速地放大到系统的输出端,变成压倒一切的差别,这种现象被称为"对初始条件的敏感性"。例如,在天气系统中,美国气象学家罗伦兹(Lorenz)首先提出了所谓"蝴蝶效应",意思是说,今天一只蝴蝶在巴西展开两只翅膀拍动一下空气,就足以使美国德克萨斯州引起一场龙卷风。这就是一种对初始条件的敏感依赖性,这一现象其实就是混沌。

混沌是由确定的过程中所产生出来的随机运动,它绝不是简单的无序,更像不具备明显的周期性和其他明显对称性特征的有序态。在理想情况下,混沌具有无穷的内部结构,只要有足够精密的观察手段,就可以在混沌之间发现周期或准周期运动以及在更小尺度上重复出现的混沌运动。

大卫·罗尔(David Ruelle)是法国的一位物理学家和数学家。因为流体流动的数学方程是非线性偏微分方程,一般无解,只在特殊情况下才有解,这是对湍流难于分析的主要原因。为此,罗尔巧妙地设想了一个抽象的图象来代替正统的俄国数学家朗道(Landau)的图象,并以美国数学家塞缪尔(Samuel)的语言来表述,把空间抽象成一个柔韧的物体,可以压缩、拉伸和折叠成马蹄形。罗尔和荷兰数学家塔肯斯(Takens)合写了一篇论文,并于 1971 年发表,其中的观点、评论和物理学理论使论文成了一篇永久的杰作。文中最著名的是作者称之为"奇怪吸引子"的图象。我们知道,相空间是我们用描述系统的独立运动变量构成的空间,相空间存在奇怪吸引子是现代科学最有力的发现之一,相空间提供了把数学变成图象的方法,把来自机械的或流体的运动部分组成的系统每一点基本信息抽象化,并绘制一条灵活的轨迹,列出每种可能性。物理学家已经得出两种简单的"吸引子"——不动点和极限环,表示达到稳定状态或连续重复的行为。取相空间任何一点作为出发点,随着时间的变化,系统的运动在相空间上描绘的轨迹最终会聚于一定的流形上。如聚在一个点上,则这个点称为一个吸引子,它表示系统的一种稳定的空间结构;若系统运动的轨迹逼近一个环面上,则称这个环为极限环,它对应系统的周期振荡。所以系统实际上只有两种吸引子:一种是极限环,另一种是不动点。每种吸引子都有自己的区域,就像两个相近的河流各有各的领域一样。以上两种吸引子都叫做平庸吸引子,而混沌则与这两者都不同。在相空间的任一起点,系统的运动都会被吸引到某一特定的区域里并将永远在这个区域中停留,但轨线不在一定的流形上而表现出不

规则的运动，称为奇怪吸引子。这时相当于系统失稳而进入奇怪吸引子。现在人们发现，耗散结构的演化，趋向稳定的平庸吸引子的可能性不大，更有可能出现的是奇怪吸引子。系统运动的轨线在相空间中的几何形态是用分数维几何进行描述的。奇怪吸引子分数维的大小在一定程度上表现了混沌行为的复杂性。

分数维几何学是法国数学家曼德尔布罗特(Mandelbrot)于 1975 年创立的一种新的几何学，以描述那些参差不齐、四分五裂的物体的图形。具有分数维几何特性的图形，其最奇妙的性质是“自相似性”，即部分与整体的形状相似，局部的局部也与整体相似。例如，一段海岸线，用放大的比例尺画出来，与整体海岸线相似；树的局部与整体的形状相似；同样，复杂的运动(指混沌)若用适当的坐标画出其图形，则其轨道构成的图形也有这种自相似性。自相似性是自然界的普遍规律。我们生活在一个三维世界之中，三维需要用三个数字才能确定一个点。空间是三维的，平面是二维的，线条是一维的，而点则是零维的，这种维数只能取正整数，是拓扑意义下的维数。曼德尔布罗特则在 0，1，2，…之外，提出了令人难以置信的分数维。分数维几何学在实践中被证明是特别有用的。分数维成为一种新方法，用以量度舍此就无法定义的客体的性质：粗糙、破碎或不规则。总之，“分形”终于成为一种新方法，可以用于描绘计算和思考那些不规则的、凹凸不平的、零散分布的、支离破碎的图形。分数维曲线代表隐藏在那些令人望而生畏的复杂图形中的有序结构，它已成为混沌学的一部分。如被称为奇怪吸引子的几何体，已被测定它具有分数维 2.06。

对混沌运动的研究有助于人们对自组织系统演化规律的认识。哈肯等人认为，无论是人造系统还是自然系统，当外界的控制参量不断改变时，大都会经历一个从无序到有序、从有序到混沌的演化序列。然而，我们必须清醒地认识到，混沌毕竟是极其复杂的现象，混沌理论发展到今天，只有二三十年的历史，目前，它还是一个年轻的、吸引人们去积极探索的领域。虽然混沌理论还不成熟，但它无疑为人们开拓了一个崭新的研究领域，激励人们用更新的目光、在更深的层次上去探索自然的奥秘。□

新兴边缘学科：多学科交叉渗透的科学

1 仿生学及其发展趋势*

仿生学(Bionics)是模仿生物来建造先进技术设备或装置的一门新兴学科。它研究生物的结构和功能的工作原理，并将这些原理移植于工程技术中，发明性能优异的仪器、装置、机器，以及为创造新的技术设备、建筑结构和新的工艺提供原理、设计思想或规划草图。简要地说，仿生学是一种"模仿科学"，是工程技术装置对有生命物质的一种模仿。

其实，人对生物的模仿古已有之。例如，我们的祖先早在 4 000 多年前就"见飞蓬转而知为车"，意即看到随风旋转的飞蓬草而发明了轮子并做成装有轮子的车。又如，人们模仿鱼的形体造船，以木桨仿鳍，以橹仿尾。

现代科学技术中也有很多是应用仿生学原理的。例如潜水服中的蹼，就是仿照青蛙的后肢做成的；潜水艇的设计，其原理也与鱼鳔的原理相同。

尽管人们很早就利用了仿生学知识，但是仿生学作为一门独立的学科并异军突起，却是 20 世纪 60 年代以后的事。1960 年 9 月美国空军航空局在俄亥俄州空军基地召开了第一次仿生学讨论会，讨论了由生物系统所得到的概念能否用于人工制造的信息加工系统的问题，即生物学能否与工程技术科学相结合的问题，并把这一新学科命名为"Bionics"。人们把这个时间定为"Bionics"正式诞生之日。1963 年，我国科学家将"Bionics"译为仿生学。

自然界的生物经过亿万年的进化与自然选择，形成了

---

* 本文完成于 1992 年。

许多卓有成效的能量转换和信息加工系统，具备了极其精确和完善的适应内外环境变化的能力。许多生物以它经济而完善的结构，可靠而协调的功能，充分显示了机器所不能比拟的优越性。于是许多工程技术人员就努力向生物系统去寻求新的设计思想和原理。这样，就在工程技术和生物学之间产生了一门新的边缘学科——仿生学。我国已故著名数学家华罗庚曾说过："生物学中有许多数学问题"，"生物的反应用数学加以描述成为工程控制论中'反馈'的源泉，神经作用的数学研究为控制论和信息论提供了现实的原型"。如蜜蜂的蜂房是由成千上万个规整的六角形巢房有规则地排列而成，它是以蜜蜂的身长和腰围为准，以用料最省为目标的一种优化了的"建筑结构"。从数学上讲，这是一个不折不扣的最优化问题。华罗庚生前对蜂房做过细致的研究，他赞扬说："巢房的精巧构造十分符合需要，如果一个人看到巢房而不备加赞扬，那他一定是一个糊涂虫。"蜂房结构原理的发现，除了进一步使最优化原理得到实际证实之外，更重要的是人们将这一原理应用于实际工程而创立了新型的"蜂窝结构"。这种蜂窝结构被公认为是重量最轻、刚度和强度最大的理想结构。由英国马丁公司最先研究出来的这种蜂窝结构，被广泛应用于航空航天工程中去。

仿生学之所以在 60 年代异军突起，是由于现代科学技术发展的需要。要使一个复杂的设备连续运动并且少出或不出故障，就必须提高它们的可靠性，而许多生物原型在可靠性方面极为突出，并且坚实、轻巧、耗能少、灵敏度高。电子装置与计算机的出现与发展，也使对生物系统的模仿成为可能。从科学发展的内在规律看，仿生学正是在知识日趋整合的基础上产生的。

生物不仅与外界进行物质与能量的交换，而且也不断进行信息的交换。比如，生物为了维持自身的生存，必须随时对外界刺激做出反应。从信息的角度来说，就是不断地同外界进行信息交换。一切生物概莫能外。许多生物信息检测和处理的能力是很奇异很高超的。金鸡报晓，青蛙冬眠，大雁秋去春来，说明生物的"时间观念"是很强的。生物这种测量时间的本领，人们称之为"生物钟"。它能使生物与自然界的时间信号保持同步。生物钟对于协调内环境适应外环境、保证有机体的正常活动具有十分重要的作用。有些生物的定位本领很高超：蝙蝠能靠超声波定位，确认目标和距离，因此它能在昏暗的夜晚飞行；海中的水母能听到空气和波浪摩擦产生的次声波，迅速地离开海岸，准确地预报即将来临的海上风暴；电鳗的头上长着一个小凸起，是它的无线电定位天线，尾端可以发射电磁波，它能够通过自身的生物"雷达系统"探明进入作用范围的各种食物，并且能够随时发出高压电脉冲击毙捕获对象等等。

生物的奇异、高超功能及其结构为人们提供了模仿的原型。如模仿青蛙眼睛制成的电子蛙眼可以用来识别飞行的导弹；人造卫星的自反差跟踪系统就是

模仿蛙眼制成的；采用电子鸽眼改进的雷达系统，可以设置在机场边缘或国境线上，对来袭导弹进行预警。

但是，仿生学也不是单纯的、机械的模仿。如飞机的发明固然是受到飞鸟的启迪，借鉴了鸟类飞行的某些原理开始的，但人们在实现“展翅高飞”的理想时，走的就不是机械仿鸟的“扑翼机”的道路，而是根据人造机器的特点，根据动力学原理，采用螺旋桨推动的“飞行机”的道路。可想而知，如果沿着仿鸟翅膀飞行的方向走下去，就不会造出与鸟飞行方式完全不同的现代飞机，更不会有遨游太空的宇宙飞船。人们在长期的历史发展中，制造工具时对生物界总是有所仿，也有所不仿，即仿创结合，仿中有创，创中有仿。

仿生的对象不拘泥于动物，还广泛涉及到植物。

一些植物学家观察到某些植物的生长是按黄金分割的序列排列的。当一株嫩芽抽枝吐叶时，如果我们从这嫩枝的顶端看下去，常可以看到其叶子的排列成一对数螺旋线，而相邻两叶在螺旋线上的距离竟符合“黄金分割”的规律。植物形态的黄金分割现象，还可在水稻、小麦等禾本科植物的茎节上看到。从第一伸长节到穗颈节，节段长度成比例增加，相邻两节段长度比例约为 1：1.618（依品种不同而异）。这正是（或接近）数学上的黄金分割比例。无独有偶，在向日葵的花盘中，种子也是按特定的对数螺旋线的弧排列的。

对所有这些曲线的研究，在工程技术和仿生学上都起到了不小的作用。例如，已弄清了玫瑰线与螺旋线之间有着极密切的关系。玫瑰线是某些机械的点描绘的轨迹，对数螺旋线与其所有向量半径成同样角度相切。如果旋转的切削刀沿这条曲线的弧运动，则可确定恒定的切削角。锄草机上的切刀就是按对数螺旋线的原理来设计的。人们根据车前草叶形排列的数学模式设计的螺旋线形的高楼大厦，使每一个房间都能得到充足的阳光。还有人模仿莲叶的皱褶结构，建造了轻巧牢固、宽敞明亮的展览大厅；利用羽茅草的叶形，搭起了筒形叶桥。至于模仿植物茎干竖起的一座座插入云霄的高层建筑，更是数不胜数了。

仿生学研究的内容非常广泛，主要有：

（1）信息仿生。它包括感官仿生、细胞和细胞间通讯仿生、动物间通讯仿生以及智能仿生等。例如，海豚能发出超声波探测前方的物体，其分辨率超过最好的声呐；响尾蛇在几米远处能感知老鼠发出的远红外光谱；狗能分辨出几千种气味。又如，令人讨厌的苍蝇和宏伟的航天事业，乍看起来，似乎是风马牛不相及，但仿生学把它们联系在一起了。苍蝇有惊人的嗅觉，能发现很远距离外的微乎其微的气味。模仿苍蝇嗅觉器官制成的高灵敏度气体分析仪，已用于分析宇宙飞船座舱内的气体成分。同时，它也可以用来测量潜水艇和矿井里的有毒气体。虽然动物的感觉器官在感知范围及定量方面有一定局限性，但综合性

能要比目前任何设计最精良的同样功能的传感器系统优越得多。因此，在工程科学的进一步发展中，人们要向生物界寻求启发并进行模拟。

（2）控制仿生。包括体内温态调控、肢体运动控制、动物的定向导航、生态系统的涨落等。如蜜蜂具有天然“太阳罗盘”，人们模仿它制成了导航设备偏光罗盘。又如，借助超声波回声定位，海豚在3千米以外便能发现鱼类，蝙蝠在飞行中能分辨直径0.1毫米的线。这些功能对提高雷达的灵敏度与抗干扰性能极有参考价值。仿生学与控制论有着共同的特点，都是将机器与生物进行类比。仿生学研究的是生物系统的结构性质、能量转换和信息过程，并将所获得的知识用来改造现有的或者创造新的机械、仪器或工艺过程，从而达到模拟人与动物某些感觉或控制功能的目的。

（3）化学仿生。它主要研究生物体内一些特殊的化学过程，以便使这些过程也能在工业中实现。如生物体中对物质合成起关键作用的生物酶，比起化学工业中使用的催化剂效率要高得多，而且不需要高压、高温等条件。生物酶的模拟成功，必将给化学工业带来一场革命。

（4）力学仿生。它是关于动物飞行动力学的研究。如按照海豚体形和身体各部比例建造的新型核潜艇，速度提高了25％。

（5）医学仿生。它主要是研究用人工仿制品代替人的有机体的某一部分。人们正在研制“人造眼”，即用光传感器模仿人眼的某些功能。这种人造眼安装在自动车上，自动车就能自动发现障碍物，改变行进方向；安装在航天器上，航天器到达目的地时将会自动选择合适的着陆地点。用压力传感器模仿人耳做成“人工耳”，可以听到连人耳也难以听到的极其微弱的声音，甚至连仓库中小虫咬吃粮食和建筑物中白蚁吞噬木头的声音，也不能逃过它的监听。

（6）机器人仿生。广义地说，机器人也是仿生的一种。现在的机器人大多是一种装有微电脑的具有程序控制能力的多功能的操纵器，它们有的像人，有的并不像人，说它们是机器人，主要是因为它们具有人的某些功能，比如眼、脑、手等器官的部分功能，把观测、思维和操作巧妙地结合起来。另外，高温、高压、粉尘、有毒、放射性等恶劣危险的生产环境下的手工操作，尤其是在宇宙和海洋开发中就更需要类似人的操作功能的自动装置替代人从事各种操作，因而产生了所谓机器人。机器人在减轻劳动强度、提高生产效率、改善劳动条件、改进产品质量、降低生产成本等方面发挥重大作用，受到人们越来越大的重视。现在已经应用和正在研究的机器人有三种：一种是工业机器人，一般具有一只手臂和存储装置，可按预先存储的程序重复地进行操作，主要在汽车制造、电子、机械等行业从事焊接、喷漆、装配、包装、零件加工以及搬运等工作；另一种是遥控机器人，其多在恶劣或危险的环境下作业，由人在安全的地方进行远距离操纵，

这类机器人已经在登月考察、火星探测、海洋开发以及核电站管理等方面得到应用；还有一种是智能机器人，它具有感觉和识别功能，能够自行决策和行动。它先是识别人给它的命令，然后根据掌握的情况，作出判断决策并采取适应外界情况的行动。现在，智能机器人已从最初识别简单的积木，发展到环境识别、规划决策、通过语言与人对话来完成某种动作的阶段，并且已从实验室走向实际应用。机器人的使用可以使生产过程逐步实现自动化。

仿生电子学是仿生学的核心，其最主要的部分是大脑及感觉器官的仿制，这也是计算机科学及机器人科学的重任。对大脑的仿制已经提出了神经网络系统的方案，并且也已设计出一些模型及实体，但由于人们对大脑的认识还不够深入，在制作方面又遇到了一时无法克服的困难，所以离完全或大部分实现大脑的功能还很遥远。

仿生学的形成和发展离不开数学。仿生学通常需要将生物系统的某些结构和功能特征进行分析，抽象出它们之间的内在联系，建立数学模型进行求解并应用。仿生学的发展与数学在生物学中的应用是相伴而行的。今天，随着数学在生物学中的广泛应用，又出现了一门新的边缘学科——生物数学。由于生命现象极其复杂，生物数学需要进行大量运算，因此这种生物数学实际上无法求解。1946 年电子计算机的问世和算法的不断改进，使得许多生物数学问题的求解成为可能，把数学在生物学中的应用推向了一个新的阶段。生物数学的出现开始了数学从非生命领域向生命领域的过渡，为用仿生学方法模拟人的智能提供了工具。

仿生学是一门方兴未艾的科学技术，有着广阔的发展前景。各行各业、各个学科都可以从中得到启发，对生物控制论无疑也是起着很大的促进作用(但仿生学还不能完全归属于生物控制论，不过同生物控制论密切相关)。尽管仿生学目前还处于成长阶段，然而它的发展必将给现代生产和科学技术带来重大突破，成为科学技术革命中众多新学科里的一支主力军。可以预言，未来许多自动化装置将会模拟生物系统的特征建立起来。各个分支的根本任务就是在本领域内争取达到对生物的最佳模拟。生物模拟将成为现代化新技术研发的重要途径之一。□

2 系统工程*

一 什么是系统工程

系统工程是20世纪60年代开始兴起的一门边缘科学，它所研究的对象是系统，特别是大系统和超大系统。所谓系统，是由两个或两个以上既相互区别又相互作用和依赖的元素（或部分）结合成的具有一定功能和目的的有机整体。一件产品、一条生产线以至一个科研课题都是大小不等、繁简不一的系统。每一个系统都包含着若干个子系统，子系统还可以再分，以至最小的单体。至于大系统或超大系统，则是指规模庞大、因素众多、结构复杂的各种工程或非工程的大型系统。如综合自动化的大型钢铁联合企业系统、空间技术（如导弹、卫星、飞船、航天飞机）的控制系统等。一般系统、大系统和超大系统之间并没有严格的分界线，在理论上这是一个模糊性问题。

1957年10月，前苏联发射的第一颗直径只有58厘米的人造地球卫星，像一声巨雷，震撼了整个美国，就连美国一般公民也感到美、苏导弹差距之大。在1961年4月，前苏联又发射了一艘乘坐宇航员加加林的“东方一号”飞船，使人类成功地进行了首次宇宙飞行。在这种咄咄逼人的形势下，1961年5月，上任伊始的美国总统肯尼迪郑重宣布，在60年代必须全力以赴地实现把人送往月球并安全

* 本文原载于《自动化博览》1992年第2期。

返回地面的目标，这就是著名的阿波罗载人飞船计划。1963年11月，肯尼迪遇刺身亡，阿波罗计划由后任总统约翰逊全盘继承下来并认真执行，直到下届总统尼克松时代，终于在1969年7月16日通过“阿波罗11号”飞船把三名宇航员送上了月球。在顺利完成科学考察任务后，飞船于7月24日安全返回地面。这就是一项杰出的系统工程，它是现代科学各个领域密切协作的结果，是系统工程方法的典型运用。为按期完成这一工程，美国当局调动了2万多家公司、120所大学和实验室的42万多名工程技术人员，为其承制了700多万个零部件，历时11年，耗资300多亿美元，实现了人类登临“月宫”探求奥秘的梦想。如此空前庞大而复杂的工程，如果不是采用系统工程的方法，使大量的组织管理技术得到协调平衡，倘若有一个环节发生故障或延时完成，都会使飞船无法按期完成计划。因此，美国的评论界对此作出权威性的论断：“阿波罗11号”登月成功，是美国系统工程的胜利，从而宣布了系统工程的诞生。

系统工程作为一门科学，还正处在发展和逐步完善的初期。有关系统工程的定义，众说纷纭，从不同的角度有着不同的理解。美国切斯纳特(Chestnut)在其1967年出版的《系统工程的方法》一书中说：“系统工程是为了研究由多个子系统构成的整体系统的具有多种不同目标的相互协调，以期达到系统功能的最优化，并最大限度地发挥系统组成部分的能力而发展起来的一门科学。”对我国系统工程有卓越贡献的科学家钱学森在其1978年出版的《组织管理的技术——系统工程》一书中说：“系统工程是组织管理系统的规划、研究、设计、制造、试验和使用的科学方法，是一种对所有系统都具有普遍意义的科学方法。”1981年他又把系统工程解释为“用定量化的系统方法处理大型复杂系统的问题”，“服务于系统的组织建立或系统的经营管理，就可统统看成是系统工程”。类似的定义或解释还有很多。总之，系统工程是以系统，特别是以大系统、超大系统为具体对象的一门跨学科的、综合性边缘科学，它是立足于总体，同时以自然科学和社会科学的理论、方法为基础，运用现代数学方法和包括计算机在内的先进科学技术为工具，对系统的组成要素、组织结构、信息交换和自动控制等功能进行分析研究、设计、制造、试验和使用，从而达到最优设计、最优控制和最优管理的目标，使局部和整体之间的关系协调配合，以实现系统的综合最优化结果的一种工程方法理论。也有人称系统工程为“科学的科学”或“工程的工程”。

系统工程属工程技术，这是多数学者一致的观点，但又不是一般的工程技术，因为一般的工程技术都有其特定的工程物质为其对象，研究的是物理。如机械工程是以牛顿的经典力学为基础，即以物质世界的运动规律的描述和应用为基础的学科。而系统工程的对象则不限于某个领域，任何物质的或非物质的

系统，都可成为系统工程的对象。它研究的是任何系统中事物的规律，研究的是“事理”，即“办事”的科学，所以有人称它为“事理科学”。系统工程是包括许多门技术的一个总称，它从横向把自然科学、一般工程技术和社会科学联系起来了。

二 系统工程方法

尽管系统工程本身的理论还很不完善，但在国际上已公认为它是一门先进的科学方法，许多工业先进的国家已开始大规模地进行系统工程的研究与应用。近十年来，系统工程的理论与方法在我国也得到了迅速的发展。钱学森就是一位杰出的系统工程专家，是我国系统工程的奠基人。1989 年，国际理工研究所为他颁发了小罗克韦尔奖章，以表彰他对中国系统工程、火箭、导弹技术、航天技术所做的重大的、开拓性的贡献。在钱学森教授等学者的倡导下，系统工程方法在我国重大的综合性项目中起到了很大的作用。如安徽淮河流域的经济开发工作、军事上反坦克武器系统的研制以及第十一届亚运会的筹备工作等，都应用了系统工程方法。所谓系统工程的方法，就是把对象作为整体系统来考虑、分析、研究、设计、制造和使用的基本思想方法，它具有以下特点：① 研究系统方法上的整体化。把研究对象作为一个整体，即把系统作为若干个子系统有机结合的整体来设计，对每个子系统的技术要求都首先从实现整个系统技术协调的观点来考虑；同时把研究过程也看做一个整体，即分析整个过程是由哪些工作环节所组成的，而后进一步分析各个工作环节的信息以及信息的传递和反馈关系，把全过程联结成一个整体，全面地考虑和改善整个工作过程，以实现综合最优化。② 技术应用上的综合化。所谓技术的综合运用，就是从各系统的总目标出发，将各种有关的技术协调配合，综合运用，以期达到整体系统的最优化。③ 管理科学化。对工程的规划、组织、控制工程进度，对各种方案进行分析、比较和决策，评定方案的技术经济效果等，统称为科学的管理。只有实行科学的管理，才能充分地发挥技术的效能。因此，管理科学化成为系统工程很重要的一个研究方面。④ 计算机化。系统工程和电子计算机形影不离，无论是设计还是管理或控制都需要使用计算机。例如综合最优化问题，遇到的变量成千上万，约束条件也是大量的，离开了计算机就无异于纸上谈兵。又如实现最优化的仿真，没有计算机的辅助也是不可能的。

系统工程方法在系统研究中的具体贯彻可简单分为以下三个互相衔接的步骤：① 模拟。模拟就是模仿、仿效的意思。我们在分析一个系统时，总是首先

建立它的模型，然后再利用这个模型对系统进行研究、试验。然而，某些事物不可能在实践之前先进行试验，有的虽能试验，但却带有很大的危险性，或试验时要消耗大量资财。这样，人们就将复杂的因素进行各种综合，来仿效事物的具体情况，这就是模拟。例如，我们不可能试打一次战争，来看看其结果如何，于是便用“沙盘地图对阵”，对战争加以模拟；在建筑工程中，要设计一座建筑物，先作出一个模型来加以研究。这些都是实物模拟。当系统工程的因素很复杂时，模拟的对象不限于实物。如设计火箭或原子能反应堆，用实物来试验不仅很危险，而且代价也是高昂的，若稍有差错，就可能导致成千上亿元的损失。另外，社会系统、生态系统等也都无法用实物来模拟。因此，系统工程中广泛采用的是数学模拟，也就是首先要弄清被模拟事物的工作原理及内部规律，用数学表达式、图形或表格等方式，定量地确定其因果关系。把所得到的这一系统列出因果关系，称为数学模型。对于一个复杂的系统，一般是通过数学模型进行分析。但这种模型能否确切地反映被模拟的对象，最终还要把它变换为能使用计算机处理的形式，以进一步进行验证，反复修正，直到满意为止。② 最优化。从数学上讲，就是在规定的约束条件下，求目标函数的最大或最小值。因此，最优化方法是一种数学方法而不是工程方法。对系统工程来说，就是使系统达到目的的最好方法，主要是解决系统的最优设计、最优控制和最优管理问题。不言而喻，这是系统工程的中心课题。例如阿波罗计划中，通常把燃料消耗的最少轨道定为最优轨道，然而，在最优条件下，从发射到抵达月球的整个飞行过程中，还要进行各种轨道修正，这即为飞行阶段的最优化。③ 系统的评价。就是利用模型和各种资料，从技术和经济两个方面对所设计的各种系统方案进行评价，权衡各方案的利弊得失，从系统整体观点出发，通过综合分析和评价，从中选择技术上先进、经济上合算、研制周期短、能协调运转而且又可能实现的最优方案。

三　系统工程的主要理论基础

系统工程是运用系统科学的观点，在控制论、信息论、运筹学及计算机科学和管理科学的基础上发展起来的，但作为组织管理的技术，系统工程的主要理论基础是运筹学。运筹学起源于第二次世界大战，当时德国拥有 1 500 多架轰炸机和 1 300 多架战斗机，而英国只有 450 架轰炸机和 600 架战斗机，英伦三岛任何一地距离欧洲大陆海岸都不超过 100 千米，德机只要 17 分钟即可到达。德国先对英国本土实行狂轰滥炸，然后实行“海狮计划”，强行登陆，最后占领英

国。虽然战前英国已发明了一套先进的雷达指挥系统，但因没有信息的快速传递、处理和显示装置，所以几次防空演习中，雷达装置虽然探测到160千米远处的敌机，但也无法将探测到的信息及时准确地提供给指挥人员使用。在发现来犯敌机的17分钟内，无法完成预警、起飞、爬高、拦截等一连串的动作，更无把握击落敌机。事实已经证明，高射炮和雷达之间的配合具有决定性的意义。为此，英国当局成立了一个由军事家、物理学家和数学家等11人参加的作战分析小组，简称“OR”小组（运筹小组）。这一小组顺利解决了高射炮指挥和雷达之间的信息传递问题，使英军挡住了敌机的强攻，成功地把敌机拦截在英国本土之外。第二次世界大战后，英、美为尽快恢复和发展饱受战争创伤的经济，把战时的经验用在经济建设上，作战研究开始形成了一门学问。传到我国后，科学家们于1964年定名为“运筹学”。“运筹”一词出自我国《汉书·高帝纪》：“夫运筹帷幄之中，决胜于千里之外，我不如子房。”运筹学研究关于人力、物力的运用、筹划和决策，以期发挥最大的效率，是系统工程最优化的主要工具。因为运筹学中的求优是从整体的角度出发，就是立足于总体概念，寻求从全局、从整体的各部分相互联系中的最优，而不是孤立地考察某个或某些局部；其次，运筹学集中了不同学科、不同专业的长处，以多学科的协同配合来解决问题，这就可避免片面性，而这些正是系统工程的要求和需要。

三十多年来，运筹学在广泛应用中形成了许多新的分支，如规划论（包括线性规划、非线性规划、整数规划、动态规划等）、排队论、对策论、决策论、存储论、搜索论、可靠性理论等。从根本上讲，这些分支无一不是求优的。随着计算机科学及算法理论的发展，运筹学的内容也在不断更新，如解线性规划的算法，已由指数时间的单纯形法发展到新的多项式时间算法——卡玛卡算法，后者比单纯形法快50倍以上，为求解大规模线性规划问题提供了比单纯形法更为有效的工具。□

3 机电一体化及其发展趋势*

在生产或服务中，应用动力驱动的机器来代替或减轻人的体力劳动，称为机械化；在没有或很少人的直接参加下，采用全部或部分能自动控制的机械设备或系统来进行生产或服务，称为自动化。1950年以后，控制技术和计算机技术的不断进步使自动化达到了新的高度。现在它不但使人摆脱繁重的体力劳动，而且能部分代替人的脑力劳动，并能完成人们原所不能完成的某些工作，故自动化是机械化的更高阶段。

机电一体化（Mechatronics）是机械技术、电子技术（特别是微电子技术）、计算机（特别是微型计算机）技术和自动控制技术的有机结合，即从系统观点出发，应用机械、电子、计算机、信息等技术，在信息论、控制论、系统论等系统科学、系统工程等基础上建立起来的一门边缘性科学。所以，它绝不是机械与电子等技术的简单拼凑，而应该看到它是一种全新技术的发展趋势。为了达到整个系统的最优化，通过机械与电子等技术的结合，互相渗透，从而产生出一大批功能更强、性能更好的新一代的机械产品和系统。如具有精度高、性能好、体积小、重量轻、自动化和智能化程度高、可靠耐用的机械产品、装置以及柔性（即灵活多样性）制造系统和自动化无人工厂等。这些机械装备和产品，往往是结构更简单、操作更方便，具有明显的经济效益。例如，数控机床、工业机器人、电子缝纫机、全自动电子照相机、全自动电子手表以及装有微电脑的各种家用电

* 本文完成于1997年。

器等等。

在工程技术和科学的发展过程中，自动化和控制技术举足轻重，它除了在宇宙飞船、导弹制导和飞机驾驶系统等领域中获得广泛的应用外，在国民经济各部门和国防工业装备中也起着重要的作用，目前它已成为现代机器制造业和电子化机器产品中十分重要而不可缺少的组成部分。

20 世纪 60 年代以来，由于生产向综合自动化方向发展，对控制设备和控制方式提出了新的要求。电子计算机的发展，特别是微型计算机的广泛应用，标志着工业生产自动化的又一次重大技术革命。80 年代，生产或系统进入综合自动化的时期，出现了便于产品更新换代、提高市场竞争能力、适合中小批量生产的柔性制造系统与机电一体化产品。可见机电一体化技术实际上是自动化技术发展到一定阶段的必然产物，它是在自动化领域中机械技术与电子技术有机结合而产生的新技术。机电一体化一词最早是在 70 年代日本使用的一个复合词，这个词的前半部分“mecha”表示 mechanics（机械学），后半部分“tronics”表示 electronics（电子学），因此，从字面上讲，应该是机械电子学，我国学者习惯地称之为机电一体化。

机电结合最初的主要方式是在机械中增加动力部件。从 50 年代初起，由于电子技术的发展，在机械中注入了前所未有的新技术，把电子元器件中的信息处理功能结合到机械中来，从而开始了机电一体化的新纪元。近年来，由于受到大规模集成电路技术的进步和普及的影响（其中特别是微型计算机），使机械产品的功能和制造技术迅速提高到一个新的水平，机械产品的构成也发生了巨大的变化，从而迎来了机械工业发展的新时期，国外也有人称之为“机电一体化革命”。

由机械技术和电子技术相互交叉、渗透形成的机电一体化技术，首先表现在机械的电子化方面。原有的机械产品由于采用了电子技术，不仅改变了产品的面貌，简化了结构，达到体积减小、重量减轻、可靠性提高、性价比改善的目的，还使机械产品具有检测控制和记忆运算处理功能，从而增加了机械的智能和灵活性。

机电一体化得以发展的根本原因，在于生产的发展和科学技术的进步，其中特别是自动化技术和计算机科学起了重要作用。第二次世界大战以后，几乎是同时诞生的系统工程、控制论和信息论这三门科学既是自动化与机电一体化的理论基础，也是机电一体化技术的方法论。而微电子技术的发展，半导体大规模集成电路制造技术的进步，则为机电一体化与自动化技术奠定了物质基础；反过来，机械制造技术也对微电子学和自动化技术做出了重大贡献。如大规模集成电路芯片的制造就是以超精密机械加工为基础的，而这种加工设备本

身又是一种计算机控制的自动化系统，即机电一体化系统。由此可见，机电一体化技术的产生既是微电子技术与自动化技术发展的结果，又是信息论、控制论和系统工程付诸实践的结果。进入80年代，机电一体化技术和产品更如雨后春笋，不断涌现，现代化的机械将电子技术、自动化技术、计算机技术融为一体，从而使机电一体化进入了所谓大发展时期。在世界范围内，各国掀起的机电一体化热潮已渗透到国民经济、社会生活的各个领域。可以说，从军事到经济，从生产到生活，从简单的消费品生产到复杂的社会生产和管理系统，机电一体化几乎达到“无孔不入”的地步。它促使产业结构、生产方式和管理体系发生了深刻的变化，促进了新兴产业的发展，同时也引起了各国为发展机电一体化控制技术的激烈竞争，从而又反过来在世界范围内更进一步推动机电一体化技术，特别是机电一体化控制技术与系统向前迅速发展。

机电一体化产品的一般工作模式是：由传感器接收外界信号，并把信号送给微型计算机；计算机对信号进行信息处理，并依据预先存入的指令、资料发出适当的命令；执行机构将计算机的命令转化为相应的动作。

分析上述工作模式可知，机电一体化的工作模式包括三个要素：① 传感器。它是外界信息的输入口，相当于人的感觉器官，如眼、耳、鼻等。它要能把机器在工作过程中机器本身和外界环境的各种参数和状况检测出来，变成一种可测定和控制的物理量，传送到机器的神经中枢。② 微型计算机。它对各种信息进行处理，即通过计算与控制实现智能功能，相当于人的头脑。③ 执行机构。当接到计算机发出的命令之后，它就去执行，具体地去实现指令所要求的动作或功能，相当于人的四肢。除去以上三个基本要素外，还应重视各种连接部分。

在微型计算机这个要素中，包括硬件和软件两部分。硬件是指微处理器、由外部总线连接在一起的存储器、输入/输出接口电路等，即计算机的机器部分。软件是指使用计算机和发挥计算机效率的一套程序系统，它包括计算机各种语言、汇编程序、编译程序、诊断程序、管理程序和操作系统等。软件是用户和计算机硬件之间的接口与桥梁，特别是操作系统是系统软件发展史上的一个里程碑。操作系统自动地管理计算机系统内部各种程序的运行，既扩充了计算机的功能，又提高了计算机的使用效率。

在机电一体化中，究竟是以电子技术为中心还是以机械技术为中心，这不能一概而论，应该依据具体的产品采取“因机制宜”的办法。譬如对数控机床、工业机器人这种类型的产品应以机械技术为中心，在现有机械结构的产品上，采用电子装置，以提高其控制性能，使之具有更好的性能或多种性能。以数控机床为例，它是在原来机床的基础上，附加上微型计算机和控制伺服机构的数控装置，用计算机的程序指令来控制机床的加工生产过程，实现加工过程的自

动化。工人只要按动键盘，输入必要的信息、参数，调用某个工件的加工程序或进行自动编程，机床就能自动进行加工。计算机数控机床的出现，使机械加工中小批量生产的自动化成为现实。目前，有些国家已停止生产普通机床，全部转为生产数控机床。如电子手表之类的产品，应以电子技术为中心，即由电子装置完全替代机械的主功能。电子手表已经用集成电路、液晶显示器和电池替代了由齿轮、弹簧等构成的精密机构。又如装有微电脑的全自动照相机，只要由人按一下快门，其余的操作便全靠存储在微电脑内的程序自动完成。最新式的电子自动照相机，除自动调节焦距、自动控制曝光外，还实现了胶卷自动装填、自动进片和自动显像，做到了五个方面的自动化。

机械与电子的有机结合，开辟出两者单独都达不到的新的应用领域。典型的例子有传真复印机、断层分析仪(CT)等。

另外，机电一体化并非局限于单个产品的范围，它也可以包括多台机电一体化设备，并由计算机控制的系统。譬如一个大型的航空航天设备，带有自动换刀装置的机床和传送装置构成的自动加工单元，甚至于一个计算机集成制造系统等，都是机电一体化的高级发展领域。

随着社会生产和科学技术的发展与进步，机电一体化正在不断地深入到各个领域，其发展前景与趋势，归纳起来，有如下三个方面：① 性能方面，向高精度、高效率、高性能、智能化的方向发展；② 功能方面，向小型化、轻型化、多功能方向发展；③ 层次方面，向系统化、复合集成化方向发展。□

4 生物数学*

生物学是研究生命现象的古老科学，在过去很长时间，生物学是应用数学最少的自然科学。恩格斯在一百多年前还说："数学在生物学中的应用等于零。"但现在情况完全改变了。由于生物学从静态的、定性的研究，发展到动态的、定量的研究，数学的理论和方法几乎渗透到了生物学的每个领域。对于生物大分子和遗传规律、生理现象和神经活动以及群体和生态的数学研究，出现了生物学和数学交叉的、崭新的一门边缘学科——生物数学。

生物系统是一个极为复杂的、性能完善的、多级结构的大系统。回顾起来，19 世纪 60 年代，奥地利遗传学家孟德尔开始用概率统计的方法研究豌豆的遗传现象。20 世纪 50 年代，数学在生物学中的应用也主要是限于初等数学。60 年代，数学理论和方法开始大规模地渗透到生物学及医学领域。例如，用微分方程描述生物链的相互制约现象；用非线性振荡模型说明生物的繁殖规律；进化论需要数理统计；种群繁衍规律依赖于概率分布等等。近十多年来，数论也在遗传密码中得到了出色的应用，就连新兴的以拓扑学为基础的突变理论，也被用以说明胚胎生长、形态发生和动物行为中的突变现象。生物或医学中还存在着许多模糊现象，因此，模糊数学在生物学及医学中也找到了它的应用。生物界绚丽多彩，生物形状极不规则，很不光滑，分数维几何学恰是描述这些气象万千的形状的有力工具。正如已故数学家华罗庚生前所写："生物学中有

* 本文原载于《现代化》1992 年第 4 期。

许多数学问题，生物的反应用数学加以描述成为工程控制论中‘反馈’的源泉。”“神经作用的数学研究为控制论和信息论提供了现实的原型”。蜂房结构原理的发现，除了进一步使优化原理得到实际的证实之外，更重要的是人们将这一原理应用于工程，从而创造了新型的“蜂窝结构”。蜂窝结构被公认为是重量最轻、刚度和强度最大的理想结构，其结果是使导弹等飞行器的研制获得成功。有人估计，21 世纪将是生物数学的黄金时代，那时生物学将会取代物理学成为使用数学工具最多的科学。

由于农、牧、渔、遗传学、医学及仿生学的需要，生物数学近几年得到了迅速的发展，派生出许多分支。如生物方程、生物统计学、生物几何学、生物运筹学、生物控制论、生物信息论等，这是按其所使用的数学方法来划分的。如按照生物学的应用来划分，又包括植物数学、数量遗传学、数量生态学、数量生理学、数量分类学等等。限于篇幅，这里仅介绍一下植物数学和数量遗传学。

植物数学

早在古代，就有人注意到数学上许多封闭曲线与某些植物的叶片和花朵的形状十分相似。直至 17 世纪，法国著名数学家笛卡儿发明了坐标法，首先把数与点联系起来，进而又把方程与曲线对应起来，从而创立了解析几何。笛卡儿用坐标法研究了一簇花瓣和叶形曲线特征，列出了 $x^3+y^3=3axy$ 的方程式，这就是现代数学中有名的“笛卡儿叶线”(或者叫“叶形线”)，数学家还为它取了一个富有诗意的美名——茉莉花瓣曲线。

数学上有一个名叫斐波那契的数列：1，1，2，3，5，8，13，21，34，…显而易见，数列中任一数(项)都是其前面两数(项)的和。许多植物学家也研究过这个数列，他们观察到某些植物的花瓣数大都是(或接近)3，5，8，13，21，34，…据记载，有一位学者耐心地数过一朵重瓣芍药花，惊奇地发现它有 235 瓣，这与斐波那契数列的第 13 项即 233 甚为接近。更有趣的是，一位学者数过一朵米切尔马斯花，它刚好是 157 瓣，其中有 13 瓣与其余的 144 瓣显著不同。他认为，这表明花瓣数目是由第 7 项($a_7=13$)和第 12 项($a_{12}=144$)合成的。人们更加注意到契波那契数列与黄金分割的关系。所谓量 a 的黄金分割，就是把量 a 分成 x 和 $a-x$ 两部分，使 x 为 a 和 $a-x$ 的比例中项，即 $a:x=x:(a-x)$ 或 $x=\sqrt{a(a-x)}$。斐波那契数列相邻前后两项的比值，即 $\frac{a_n}{a_{n+1}}$(a_n 是第 n 项的数值)，

当 n 无限增大时，是以黄金分割数$\frac{\sqrt{5}-1}{2}\approx0.618$为极限值。$n$ 值越大，则该数列中前后相邻两项的比值就越接近 0.618。例如，5∶8=0.625，8∶13≈0.615，13∶21≈0.619，144÷233≈0.618 等等。

一些植物学家还观察到某些植物的生长也是按黄金分割的序列排列的。当一株嫩芽抽枝吐叶时，如果我们从这嫩枝的顶端看下去，常可以看到其叶子的排列成一对数螺旋线，而相邻两叶在螺旋线上的距离竟服从黄金分割的规律。植物形态的黄金分割现象，还在水稻、小麦等禾本科植物的茎节上看到。

植物学家也研究过植物的叶片，虽然它们的几何图案各不相同，但它们的茎和叶横切面与叶脉之间却有着最经济的对应关系。拿来一片绿叶，我们会看到其上面有纵横交错的叶脉，它们都是叶子的“运输线”，在植物学上叫做叶子“维束管”。它们能以最少量的叶脉而布满全叶的有效面积，达到最佳的运输效果，这就是植物形态结构中存在的一种最优化生活系统，它体现了数学上的最优化原理。倘若我们把这种最优结构运用到市政工程上，就能设计出简便而经济的供水或煤气管道。如果铺设管道的地区狭长，就像高粱、玉米叶子那样的形状，那么，导管系统就应接近这些植物叶脉的分布图案。由此看来，一些植物系统最优化的研究，将为工程系统提供新颖的设计思想。

数量遗传学

研究生物的传宗接代的学问叫做遗传学，而生物遗传学中的密码问题是研究遗传与变异的一个根本性问题。每处生物都有自己的“遗传密码”，也叫“遗传信息”。不论是动物、植物还是微生物，它们都把各自的遗传密码传给了下一代，下一代就按照这个密码生长发育。俗话说“种瓜得瓜，种豆得豆”，瓜里有瓜的遗传密码，豆里有豆的遗传密码，代代相传，互不干扰。这都是由生物的遗传密码所决定的。那么，什么是遗传密码？

谈到密码，有点电报常识的人都知道，中文电报是从 0，1，2，3，4，5，6，7，8，9 这 10 个数码中任选 4 个，按不同顺序编成密码，这些密码分别代表成千上万个中文汉字。用 4 个数码作为 4 个符号来编码，叫做四联密码。在电报通讯上，可以用四联密码组合成各式各样的电报内容。因为局外人还不能识别这种编码，所以叫做密码。

遗传密码和电报密码很相似。遗传学研究发现，一切生物负责传宗接代的物质都是核酸。现已研究清楚了，核酸是一种化学物质，是一种生物大分子，核

酸分子是同4种不同的核苷酸小分子连接起来组成的一种变化无穷的锁链。这4种不同的核苷酸就相当于电报中4个不同的数码,其中每3个组成一个"密码子",可以称为三联密码。各种密码子组合排列,能代表成千上万的遗传信息(实质就是基因,一个基因代表一种信息)。这种核苷酸编码叫做遗传编码。因为人们最初还不能识别这种编码,所以叫做遗传密码。由许多三联密码组成的基因好比电报里的一个中文句子,由无数基因组成的核酸分子好比一份完整的电报。因为核酸是一种生物大分子,所以一个核酸分子包含着许多的遗传密码量。据估计,可能比一部中篇小说的字数还要多,所以有人风趣地说:"人的遗传密码编写了一部世界上最巨大的著作。"

世界上的生物尽管千差万别,但所用的密码系统都是相同的,就像全国邮政系统采用的是同一个密码本一样。从60年代以来,经过各国科学家的艰苦劳动,全部密码被一一破译出来了。发现生物的密码全都是三联密码,按数学上的排列组合方法计算,总共有64(即4^3)种密码,并且把它汇编成了密码表。所有生物都是按照这同一本密码来为这各种各样的遗传性状编码的,所以说,生物的遗传性状是由不同遗传密码决定的。

由于生物大分子是一种大系统,就必然地要使用较多较深的数学理论和方法,这种采用数理统计和数学分析方法研究数量性状遗传的遗传学分支学科,就是数量遗传学。

由于生命现象十分复杂,用数学方法研究生物学,需要进行大量的数据处理和复杂的数学运算,单靠人工劳动是困难的,电子计算机和计算机科学的出现与发展,为生物数学的迅速发展提供了物质基础。如同望远镜之用于天文学和显微镜之用于生物学打开了人类认识自然的新境界一样,电子计算机为生物数学的研究开拓了新的天地。同时,生物数学的出现,标志着数学从非生命科学的研究转向生命科学的研究,也为用仿生学方法模拟人的智能提供了新的工具。□

5 遥感、遥测和遥控*

遥感(remote sensing)是1960年由美国学者普鲁依特(E. L. Pruitt)提出来的。顾名思义，遥感应该是在遥远处感知事物的意思，不过，在地面上所能观察到的"遥远"是有限的。障碍物能挡住我们的视线，即使是毫无遮挡的"一马平川"，由于地面是个曲面，我们也不可能看到太远的地方。"欲穷千里目，更上一层楼"的意思是说，登高才能望远。哪里最高呢？天上——地球的上空。然而，当今所谓的遥感仅限于电磁波遥感，它以地球为研究对象，通过航空航天器所载的电磁波传感器(如扫描照相机和辐射计)来收集地面目标所辐射或反射的电磁波信息，经过记录、传输和处理，变成人们可以直接识别的信号或图像，并从中获取有关的特征信息。因此，遥感是一门应用现代科学技术来探测有关地球信息的新技术，它利用航空航天器从高空通过电磁波(光波或无线电波)来感知大地上有关目标的特征或摄取其图像，而后进行分析处理与识别。

1957年，前苏联第一颗人造地球卫星上天，从这以后，随着空间技术、电子技术的进步，遥感技术有了质的飞跃。应该说，地球上的一切物体，都无时无刻不在向我们发出包括光和热在内的各种电磁波，每种物体发出的电磁波都有自己的特征，表明各种物体的存在方式和运动状态。遥感按所用的电磁波频段可分为可见光遥感、红外遥感和微波遥感等，不同频段有不同的使用情况。可见光遥感最常用的主要探测仪器有全景航空摄影机、多光谱照相机等多

* 本文原载于《自动化博览》1999年第2期。

种。多光谱照相机有好几个镜头,可以同时摄取同一目标发出的不同波段光线的照片,也可以说,它有好几只"眼睛"能分别"看见"不同颜色的光。根据需要,把它们进行不同的组合,就能获得地面上各种资源的图像。红外遥感的主要仪器有单通道红外辐射计、红外热图像仪以及多通道红外扫描仪。它们的主要特点是能够在夜间工作,不管是什么物体,只要是它的温度在绝对零度以上,它就会不停地辐射红外线,人眼看不见,红外探测器却能感受到它们。这就是说,一个红外探测器就是一只"夜眼睛",那怕是在伸手不见五指的黑夜,它也能"看到"地面上物体的清晰图像。红外探测器还有识别伪装的能力。比方说,坦克和汽车漆成绿色,或者用树枝把它们伪装起来,它们可以骗过人的眼睛,却骗不了红外探测器,因为绿色油漆或采下来的树枝对红外线的反射和辐射能力比活的绿色植物差得多,所以在红外探测器的"眼睛"里就显现了原形。用可见光照相机照不出地热、水体、植被等红外辐射异常所引起的变化,用红外探测器却能探测出来。微波遥感仪器有两类:一种是无源遥感器,像可见光照相机和红外探测器一样,能接收目标所辐射或反射的微波;另一种是有源遥感器,比如雷达,本身能够发出电磁波,然后接收目标反射回来的信息,显示出目标的图像来。微波能够穿云透雨,还能穿透森林和一定深度的冰雪、土壤,具有全天候工作的能力,是探测热带雨林、极地冰盖下地貌和基底构造的重要手段。

遥感还有航空遥感和航天遥感之分。前者使用飞机或气球,分辨率高,但探测范围较小;后者使用人造卫星和航天飞机,分辨率较低,但探测范围较广。

遥感技术的广泛应用,为我们提供了一种大面积、高速度获取信息的重要工具。

拿"大面积"来说,原因就在于它站得高,看得远。从以飞机为主要运载工具的航空遥感发展到以人造卫星为主要运载工具的航天遥感,高度一下子增加了好多倍,一次拍摄的面积也就大大增加了。一张航空照片拍摄的地面面积通常只有几十到几百平方千米,要把我国的领土全部拍摄一遍需要 50 万到 100 万张照片。采用航天遥感,人造卫星离地面 1 000 千米,一张照片就能照下 3 400平方千米的地面面积,相当于一个海南岛那么大。离地面更高的人造卫星或飞行器,可以把我国的领土拍摄在一张照片上。

再说"高速度"。人造卫星在太空中绕地球转圈,飞行速度极快。过去实地测量一个地区的农田或地形,得花几年甚至几十年时间。现在用人造卫星探测,一天拍摄地面照片 180 多张,18 天就可以把整个地球拍摄一遍。这就为我们研究和征服台风、洪水、火山、地震等重大自然灾害创造了条件。

遥感不仅距离遥远,而且感知灵敏。人造卫星上的照相机可以在几百千米的高空中分辨出小于一平方米的物体。遥感使我们的感官延长了,眼界开阔

了，它可以帮助我们侦察和监视各种军事设施、武装力量的部署和调动情况，发现隐蔽在森林里的车辆和坦克部队，大面积收集植物、土壤的资料，掌握地表水、地下水的分布和变化，了解农作物长势和病虫害，精确测量各种主要气象参数，提供全球性的云图照片，为大范围气象预报提供资料等等。

总而言之，遥感作为一门综合性高新技术，观测空间大，可远离被测物体，不受气候条件的限制，在世界范围内环境与灾害的监测中一直被优先选用。从80年代末开始，我国科学家即开始遥感技术在洪涝灾害监测中的应用研究。经过国家863计划和“八五”、“九五”攻关实验，进一步完善了监测技术手段，形成了卫星、飞机并用的完整体系。1998年夏季，我国长江中下游、嫩江、松花江流域遭受百年不遇的大水灾，在中央科技部的统一部署下，从6月19日起，我国遥感工作者利用6颗不同种类的卫星和3个遥感监测系统，在地理信息系统和全球定位系统的支持下，获得了大量洪水实况信息，上报给国家防总等部门，使遥感在抗洪防灾中显示了威力。

遥测是一门应用通信技术对远距离的物理量进行测量的新技术，它属于测量学的一个分支。遥控则是一门利用现代通信技术对远距离的或人不能接触的（如高温、有毒或辐射等）物体或系统的运动进行控制的新技术，它是控制学科的一个分支。遥测与遥控经常结合在一起使用，即根据遥测到的各种参量数据实施对目标系统的遥控。遥测、遥控技术是通信技术与测量、控制技术相结合的产物。

遥测技术主要是研究如何将传感器获得的测量信息快速可靠地传送到远处的目的地。遥测类似于遥感，不同之处在于传感器是放置在被测物体上或在其附近，而不是远离它。遥测系统包括各种物理量传感器、信号变换器、多路转换开关、调制器、发射器、传输通路和接收器等。按传输方式，有单通道遥测和多通道遥测之分；按通信形式，有有线电遥测和无线电遥测之分；按物理量表达方式，则有模拟式和数字式之分。其中最主要的是各种物理量的传感器。人的感觉器官（耳、眼、鼻、身等）都是传感器。人类认识外界事物的过程，是外界事物作用于人的感觉器官产生相应的响应，并把这种响应通过神经系统传递到大脑的过程。随着社会生产的发展和科学技术的进步，人们越来越感到单靠自己的感官来收集外界的信息已经不够用了，于是研制成了各种各样的传感器（如望远镜、显微镜等），以代替人的感官去感知外界事物的变化，这样，传感器又得到了“人工器官”的美称。在自动控制系统或其他装置中，传感器被用来感知或反映某一设备的工作状况，所以也被叫做测量元件。在人不能到的地方或对人有伤害、有危险的地方，可以派传感器去工作，起到人的感官的作用。

人的感官是最了不起的传感器。不过，人造的传感器有时也可以突破人的

生理界限，感受到人所不能感受到的外界信息。其实，遥感系统之所以具有“遥感”的能力，就是因为它们装备了传感器，特别是光和热的传感器。

传感器多种多样。有的像眼睛，对光特别敏感，故又称为光传感器。光传感器有很多用处，包括测光、计数、自动控制、电视、录像、电影放映等，我们在日常生活中也能遇到它。比如说，用光传感器做成楼房大门或电梯的自动开关、高级照相机的曝光机等等。它也可以用于全自动洗衣机的自动控制，根据水的透明度的大小来监视衣服是否已经洗净；衣服洗干净了，光传感器会发出信号，让洗衣机自动停车。过去光传感器只能感受光的强弱，现在新型的光传感器已能像人眼那样识别光的颜色。这个进步，就像由黑白电视机发展到彩色电视机一样。人们还在研制“人造眼”，用光传感器模仿人眼的某些功能，如接收物像，进行测量并传送信息。这种人造眼安装到自动车上，自动车就能判明障碍，改变行进方向以避免碰撞；安装在航天器上，航天器到达目的地时将会自动选择合适的着陆地点。有的传感器像耳朵，能“听”到极微弱的声音：听觉是由声波的机械振动引起的，振动又会产生压力的变化，从这个意义上说，听觉器官就是一种压力传感器。人造的压力传感器有金属电阻型、压电电阻型等几种。压电电阻型用得最多，由锗、硅等半导体材料制成，一旦受到压力，电阻率发生变化，它就把压力信号转变成电信号。应用压力传感器，可以做成“人工耳”，听到连人耳也难以听到的极其微弱的声音。人工耳的听觉非常灵敏，甚至连仓库中小虫咬吃粮食和建筑物中白蚁吞噬木头的声音，也不能逃过它的监听。用压力传感器还可以做成人工耳蜗，人工耳蜗把听到的声音变成电信号，传给埋在听觉神经末端的微电极，刺激听觉神经末端，使聋人回到有声世界中来。

压力传感器不光用作人工听觉，也可用作人工触觉。比方说，把几十个钛酸钡压电元件以 1～3 毫米的间隔排列起来，用微型计算机控制，就制成了一个人工触觉器，它体积小，重量轻，那怕只有 10 微米的凹凸不平，它也能感觉出来。把人工触觉器装在假手上，可以使手有触觉；装在机械手上，可制成有触觉的机械手，这种机械手能判断物体的大小和重量。

温度传感器能把温度变化转换成电信号。它收集温度信息的范围特别宽，从使所有气体变成液体的接近绝对零度的超低温，到能引起核聚变的几千万度的超高温。它的测量精度也特别高，可以在 －263℃～180℃ 的范围内测出 0.001℃的变化。温度传感器在收集温度信息的时候，不一定非得与被测物体相接触不可，它可以远离物体进行遥测、遥感，这就给我们带来了很大的方便。例如，列车的轴瓦在运行中很容易发热、烧损，过去铁路工人必须在停车以后逐个用手摸测，现在只要在离车站几千米以外的铁轨上安装一个温度传感器，列车从它旁边通过就能知道哪节车箱的哪一个轴瓦有问题。它能发现架空高压

线上和运行电器设备中的过热点，控制轧钢机的生产，监视导弹的发射等。

具有嗅觉的人工器官是气体传感器。人鼻嗅到的只是气体的香臭，而气体传感器却能用来分辨不同的气体。气体传感器又叫电子鼻或气敏仪，虽说它是模仿人鼻的功能，但它比人鼻的嗅觉还灵敏。比方说，它可以毫不费力地“嗅”出一氧化碳、苯等有毒气体，比人的鼻子灵敏得多。现在，用气体传感器已能分辨出三四十种易燃、易爆或有毒气体了。

值得一提的是，传感器正向集成化和多功能化的方向发展。也就是说，一个传感器可以传感几种化学或物理量的变化。比如既能感觉温度，又能感觉气体等。未来的传感器将变得越来越灵巧，具有一定的智能和判别能力，不仅模仿人的五官，而且模仿人的大脑。灵巧化的传感器既能收集各种信息，把信息变成电信号，又能对这些信息进行分析、处理，然后把它显示出来，或者输送给执行装置，去完成自动控制的任务。

遥控技术主要研究如何将控制信息快速可靠地传送到被控对象。遥控系统类似于遥测系统，也包括操作数据发送器、信号变换器、调制器、发射器、传输通道、接收器和解调器等。同样，遥控系统也有单通道和多通道、有线电与无线电以及模拟式与数字式之分。

遥测与遥控技术早期多用于电力、天然气、燃油以及自来水等供应的测量控制，现已推广应用于航空、航天和交通等领域。遥测与遥控的发展方向主要是提高测控精度、增大测控距离、加强测控的可靠性和减小测控部件的体积与重量以及开辟它们的应用领域等。□

后记:往事在昨天

范伟力

子都晚年的研究成果变成一本厚重的书,令我充满欣喜又不禁感慨万端。他若地下有知,该有何等地欣慰呀!

亡人已经走远,时间却不会湮埋生者的痛。回首往事,桩桩件件竟如此清晰,犹如昨天。儿女说,那些过去的事,就不要再提了!但我认为,有些记忆只是记忆,有些记忆就是历史,历史和现实是衔接的,历史有时是需要翻看的。

1965年,知识界开展"向党交心"活动。组织上认为子都是有影响的老同志,就动员他起个带头作用,哪怕一闪念也要向党忠实交心,并承诺执行"不做记录、不编辫子、不戴帽子、不打棍子"的"四不原则"。子都憨厚,总拿鸡毛当令箭,于是极认真地写了几页稿纸,还拿到大会上做了发言。谁知"文革"的大潮袭来时,全院第一张大字报针对的竟是他的发言稿!

随着运动的升级,子都的"资产阶级反动学术权威"的罪名升格为"历史反革命按人民内部矛盾处理",被遣送山东原籍。

少年离家35年,没有荣归故里,反倒戴罪还乡,子都愧疚又谦卑,对乡亲们的第一句话就是:"我回家参加劳动,接受贫下中农的再教育来了,请你们多操心。"大家见他如此诚恳,倒笑了,说:"在咱家乡没有再教育这事儿,人人挣工吃饭,天明干到天黑。"家乡人善良纯朴,父老乡亲没有一个嫌弃他,儿时学友纷纷赶来探望叙旧,促膝长谈。子都在家乡学习劳动处处积极,清晨上工,他比生产队长还要提前。山上割草,他挑起满满的担子,无论怎样东摇

西摆，他都咬紧牙关坚持到底。别人劝他不用那么认真，少割一点儿草，也是个好社员。

子都回老家没有带什么东西，却带了一个大木箱子，里面装的全是他的书稿讲义和各种数学书籍。他舍不得扔下它们。劳动之余，他就读书，就着微弱的煤油灯常常读到深夜。

我第一次去山东看子都，见他人黑瘦，精神有点恍惚。亲友们说他太劳累，却又舍不得到集市买一点儿营养品。我听着难过地哭了。那时子都急切地想了解上面的政策，渴望公正解决自己的问题。子都他是多么希望早日回到工作岗位啊！

1971年毛主席签发的落实知识分子政策的44号文件下达，我单位的领导亲自将文件送到我的手里。看了文件内容，我连夜抄写寄给子都。收到信件的他也激动万分，对着毛主席像连呼三声万岁！但政策在基层落实的阻力非常大，学院人事部负责人声色俱厉："我们是红色政权，确定的事情不能更改！"我不甘心，无数次奔走呼吁，直至市委组织部确认：按照中央44号文件精神，属人民内部矛盾的，可在任用中改造。

1972年寒假，我最后一次去山东探望子都，夫妻相对，泪如雨下。子都当时住在一间偏房里，屋内没有任何取暖设施，冰冷凄清，他的两耳冻出了水泡，面容憔悴，失眠和胃痛时刻折磨着他。我泣不成声地对他说："你的问题虽然没有最后落实，但也很有希望。咱们回家吧，回家治病，我能养得起孩子，就能养得起你……"

子都回来引起很大反响，有人说，胆子太大了，没落实政策就敢回来！更有甚者指着他的鼻子喝斥："你是什么人？早点回去劳动改造！"子都灰溜溜地不敢吱声，我听说后非常气愤，坚信问题总有水落石出的一天！不久，市委组织部专门派了工作组到学院办了一个"转弯子学习班"，督促落实知识分子政策。

子都终于正式上班了，此时与44号文件的下达整整间隔了两年。然而当他怀着满腔的感恩和希望去单位报到时，却被安排到实习工厂当车工，"继续接受工人阶级的再教育"。他没有提出任何异议，也没有丝毫怨言，每天老老实实去车间上班。尽管他加倍努力，怎奈驾驭车床的技术不争气，加工的螺丝多是不合格的废品。为了掌握技术，他经常加班加点干到深夜。超负荷的劳动和沉重的心理压力使子都寝食难安，经常地，吃了几口饭，就要出去走走，我只好让孩子尾随其后，生怕他想不开寻短见。只有我理解他那时的心境是怎样地苦不堪言。

1973年，学院复课，开设工农兵中专班，教师们纷纷重登讲台，唯独子都没有恢复教职。我实在不忍他这样委曲求全了，找到领导质疑："赵子都当车工和

当教师哪个更适合他？哪个贡献大?”这一次得到了应有的答复，子都很快转入教学岗位。这一年，他53岁了。

“文革”结束后，重被委以重任的子都以极大的热情投入到他终生热爱的数学与教学中，比从前更加忙碌更加执著。他被评为全省模范教师，还实现了入党的夙愿。

2000年4月18日，在子都的追悼会上，学院领导以“鞠躬尽瘁，死而后已”为他的一生划上了句号。我深以为然，因为他的品格、他的学术造诣、他的敬业精神是得到全院师生高度认同的。

大学，聚集着知识的力量，并且遵循着知识与科学的逻辑前行，违背这一根本的规律，只能制造荒谬和悲剧，只能成为讲述历史时一个醒示后人的案例。

2009年3月

[General Information]
书名=数林奇葩
作者=赵子都著
页数=308
SS号=12402430
DX号=
出版日期=2009.07
出版社=山东教育出版社